UNCERTAINTY QUANTIFICATION FROM SMALL DATA: A MULTIMODEL APPROACH

by

Jiaxin Zhang

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

August, 2018

Abstract

As a central area of computational science and engineering (CSE), uncertainty quantification (UQ) is playing an increasingly important role in computationally evaluating the performance of complex mathematical, physical and engineering systems. UQ includes the quantification, integration, and propagation of uncertainties that result from stochastic variations in the natural world as well as uncertainties created by lack of statistical data or knowledge and uncertainty in the form of mathematical models. A common situation in engineering practice is to have a limited cost or time budget for data collection and thus to end up with sparse datasets. This leads to epistemic uncertainty (lack of knowledge) along with aleatory uncertainty (inherent randomness), and a mix of these two sources of uncertainties (requiring imprecise probabilities) is a particularly challenging problem.

A novel methodology is proposed for quantifying and propagating uncertainties created by lack of data. The methodology utilizes the concepts of multimodel inference from both information-theoretic and Bayesian perspectives

to identify a set of candidate probability models and associated model probabilities that are representative of the given small dataset. Both model-form uncertainty and model parameter uncertainty are identified and estimated within the proposed methodology. Unlike the conventional method that reduces the full probabilistic description to a single probability model, the proposed methodology fully retains and propagates the total uncertainties quantified from all candidate models and their model parameters. This is achieved by identifying an optimal importance sampling density that best represents the full set of models, propagating this density and reweighting the samples drawn from the each of candidate probability model using Monte Carlo sampling. As a result, a complete probabilistic description of both aleatory and epistemic uncertainty is achieved with several orders of magnitude reduction in Monte Carlo-based computational cost.

Along with the proposed new UQ methodology, an investigation is provided to study the effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets. It is illustrated that prior probabilities have a significant influence on Bayesian multimodel UQ for small datasets and inappropriate priors may introduce biased probabilities as well as inaccurate estimators even for large datasets. When a multidimensional UQ problem is involved, a further study generalizes this novel UQ methodology to overcome the limitations of the independence assumption

by modeling the dependence structure using copula theory. The generalized approach achieves estimates for imprecise probabilities with copula dependence modeling for a composite material problem. Finally, as applications of the proposed method, an imprecise global sensitivity analysis is performed to illustrate the efficiency and effectiveness of the developed novel multimodel UQ methodology given small datasets.

The content in this dissertation has been presented in the following publications:

J. Zhang and M. D. Shields. "On the quantification and efficient propagation of imprecise probabilities resulting from small datasets." *Mechanical Systems and Signal Processing* 98 (2018): 465-483.

J. Zhang and M. D. Shields. "The effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets." *Computer Methods in Applied Mechanics and Engineering* 334 (2018): 483-506.

Primary Reader and Advisor: Michael D. Shields

Secondary Reader: Lori Graham-Brady, Sauleh Siddiqui

Acknowledgments

I would like to express my greatest thanks and sincere gratitude to my advisor, Prof. Michael D. Shields, who has been providing me with excellent guidance, extensive support and continuous encouragement throughout my Ph.D. study. He was always there to offer invaluable opportunities and suggestions when I need it. I sincerely appreciate his patience, understanding and trust when I encountered difficulties. Prof. Shields is also a role model for me - his dedication, attitude, ethics and his style of work will follow me and inspire me throughout my professional career and future life. I owe my special gratitude to him.

I would also like to thank Prof. Lori Graham-Brady and Prof. Sauleh Siddiqui for serving on my thesis committee and contribution valuable feedback and suggestion on my work. I would express my sincere appreciation to Prof. James Guest, who helped me patiently in the early period of my Ph.D. study. Specifically, in the Department of Civil Engineering, I would like to thank Prof. Somnath Ghosh and Prof. Stavros Gaitanaros for their great teaching in me-

chanics courses which consolidated my foundation in this field. Additionally, I am grateful to my colleagues at SURG. Many thanks to Hwanpyo, Sundar, Aakash, JP, Dimitris, Mohit and Lohit for many insightful discussions in research and great time that you spent with me.

My sincere thanks to Prof. Ben Hobbs, who served as my GBO committee chair, for his some advice that helped me professionally in the long-run. I would also like to express my gratitude to Prof. John Wierman, who is my Master's advisor in Applied Mathematics & Statistics, always provided his experience, suggestion and help in developing my professional path. I would like to thank Prof. Mengyang Gu and Prof. Yanxun Xu for our discussions on my research work, particularly involving Bayesian statistics. I also want to appreciate Prof. Fei Lu, who offered useful advise from the perspective of mathematics when we meet at the conference. In addition, many thanks to Prof. Antwan Clark for his discussion and help in understanding reliability analysis.

I would like to acknowledge the grant support from the Office of Naval Research under Award Number N00014-15-1-2754 and N00014-16-1-2582 with Dr. Paul Hess as the program officer. The travel support from HEMI, USACM, SIAM and Acheson J. Duncan Fund in Applied Mathematics & Statistics at JHU are also greatly acknowledged. I would like to thank Prof. Stephanie Ter-Maath at the University of Tennessee, for her assistance and discussion in the composite material model. In addition, I need to thank Prof. Ling Liu at Utah

ACKNOWLEDGMENTS

State University, who provided me with a valuable collaborated opportunity for UQ application.

Finally, I would like to express my gratitude to my parents for their continuous support. Specifically, I would like to thank my wife, Sirui who encouraged me when I was depressed and doubted myself and supported me throughout.

Thank you all!

Dedication

This thesis is dedicated to my parents and my dear wife, Sirui.

Contents

List of Tables

LIST OF TABLES

List of Figures

Chapter 1

Introduction

1.1 Overview

Uncertainty quantification (UQ) is the science of quantitatively characterizing and reducing uncertainties in the context of computational science and engineering (CSE) [4]. Practically speaking, UQ is playing an increasingly critical role in reliability analysis, risk evaluation, verification and validation (V&V), performance prediction and decision making [5]. UQ is also widely applied to many scientific and engineering fields including applied mathematics and statistics [6–8], computational physics [9–12], computational mechanics and materials [13–15], energy and environment [16, 17], etc. Uncertainty can be broadly categorized into two classes: *epistemic* uncertainty resulting from a lack of knowledge, data or imperfect models and/or assumptions, and *aleatory*

uncertainty resulting from natural random [18]. Ferson and Ginzburg [19] argued that unlike aleatory uncertainty, epistemic uncertainty needs a different mathematical framework. However, it is still open to debate: What is the most appropriate mathematical treatment for epistemic uncertainty?

1.1.1 Imprecise probability

Given the intuitive nature of probability theory, it is common to consider all uncertainty probabilistically. Even though aleatory uncertainties are often considered using standard probability theory, epistemic uncertainty often shows a level of "imprecision". A related field, termed imprecise probabilities, has been therefore proposed. Walley [20, 21] developed a unified theory of imprecise probabilities, but there are still a number of methods to investigate the imprecision, which mainly includes probabilistic and non-probabilistic theories. Probabilistic approaches include probability boxes (p-boxes) [22–24], Bayesian [25, 26], random sets [27–29], and frequentist [30–32] theories and combinations of these approaches [33, 34] among many others (e.g. [35]). Non-probabilistic methods include interval methods [36, 37], information theory [38], convex models [39], fuzzy sets [40] and Dempster-Schafer evidence theory [41, 42]. Beer et al. [43] presented an extensive review for many of these theories in engineering applications. The interested reader may find more details involving the application of imprecise probabilities in [43].

1.1.2 Uncertainty caused by lack of data

In statistical theories, Bayesian and frequentist methods perform well for problems if large datasets are available. Methods to handle so-called big data [44, 45] have been widely used in recent research and in practice, such as machine learning, deep learning and artificial intelligence (AI) [46–49]. However, it is often difficult to collect large datasets for many engineering applications. In many cases, data collection comes at very large cost. As a result, it is necessary to investigate how to improve the inference, estimation and prediction given small datasets. Small data creates a specific type of epistemic uncertainty (sometimes referred to as second-order uncertainty [50]) which introduces difficulty in identifying a unique and accurate model for the underlying probabilities. The motivation of this work therefore aims to propose a probabilistic methodology that addresses the primary challenges in UQ given small datasets [3, 51–53]: model-form uncertainty, model parameter uncertainty and uncertainty propagation.

1.1.2.1 Model-form uncertainty

In UQ applications, it is desireable to assign a specific model-form (for instance, probability distribution model) because it simplifies the uncertainty analysis and clarifies the prediction and decision making. However, it raises a question whether it is reasonable or justified to assign a unique model based

on very limited data. More commonly, a probability model is assigned according to either expert judgment, experience, or through a comparative down-selection or averaging process [54]. On one side, regarding the comparative down-selection process, a number of candidate models are first provided and then one selects the "best" model based on some criterion, such as Bayesian hypothesis testing [55], or Bayesian model selection [25], information-theoretic model selection [56], or goodness-of-fit tests (e.g. [57]). Alternatively, averaging methods assign relative weights to the models that are being considered and combine them. Sankararaman and Mahadevan [26] applied Bayesian model averaging to study the uncertainties resulting from small datasets. However, rather than selecting a single model given lack of data, we prefer a multimodel inference method based on the theory presented in [58] and more explanations can be found in [3].

1.1.2.2 Model parameter uncertainty

Given a specified model, inference on the model parameters is used to quantify uncertainties for a given model. Classical frequentist methods, such as bootstrapping [59], estimate the parameters as deterministic values with confidence bounds. This can be problematic for the case of small data. The Bayesian approach treats the model parameters as random variables and estimates the joint probability distribution through Bayes' rule. Rather than identifying the

joint parameters as a point estimate (selecting a single set of maximum likelihood estimates of model parameters [60]), this work retains the full joint parameter densities for each candidate model and therefore considers the total uncertainties when combined with the model-form uncertainties identified by multimodel inference.

1.1.2.3 Uncertainty propagation

The multimodel methodology retains the full model-form and model parameter uncertainties. But propagation of multiple probability models introduces large computational expense [50]. Sankararaman and Mahadevan [26] pointed out that propagating this uncertainty involves multiple loops of Monte Carlo simulations that come at very large computational cost. Fig. 1.1 illustrates those loops over the probability models, model parameters and Monte Carlo sampling from each selected model. In this work, a novel single-loop Monte Carlo approach is proposed to retain the full uncertainties and efficiently propagate them through a computational model, shown in Fig. 1.1(b). The proposed methodology simultaneously propagates uncertainties from a full set of candidate probability models, each having uncertain model parameters.

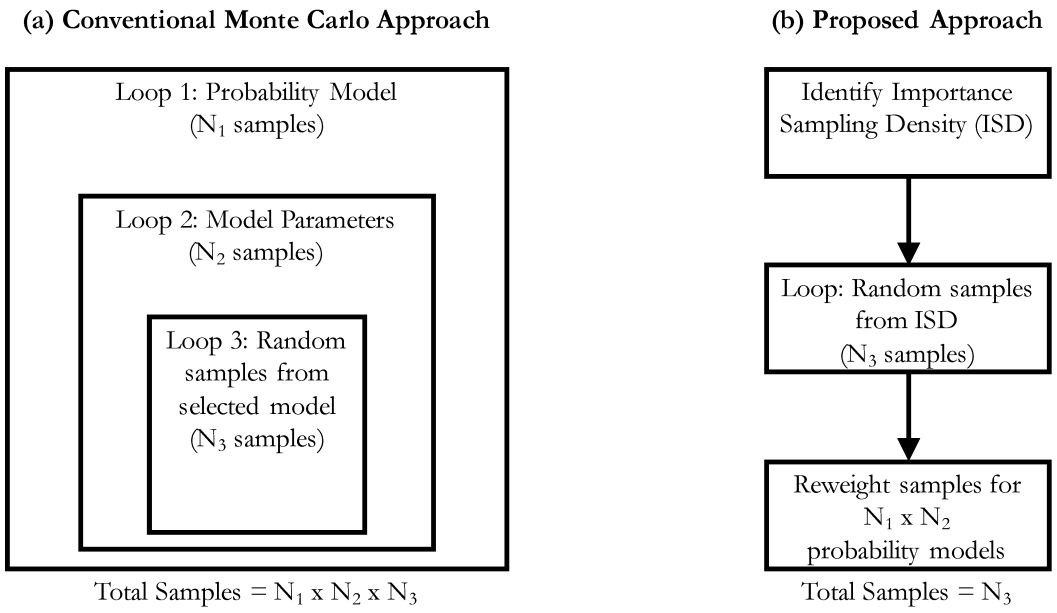


Figure 1.1: Conceptual comparison of (a) the conventional Monte Carlo approach for propagating model-form and model parameter uncertainty, and (b) the proposed approach (image from [3]).

1.1.3 Effect of prior probabilities on UQ

Considering the approaches employed herein use Bayesian inference from limited data, prior information may influence the uncertainty quantification and propagation in important ways. One primary object of this work is to enhance the understanding of the effect of prior probabilities in this Bayesian multimodel uncertainty quantification framework [61]. Commonly, noninformative priors are used for estimation of model parameter uncertainty using Bayesian inference. Given lack of data, informative prior knowledge may improve the inference and prediction of model-form and model parameter uncer-

tainties. But, sometimes, the informative prior may be inaccurate in seemingly subtle ways such that it leads to a biased or even wrong estimate. This work aims to systematically discuss the performance of noninformative and informative priors and compare their effect on Bayesian multimodel uncertainty quantification and propagation.

1.1.4 Correlation/Dependence modeling in UQ

Many UQ approaches assume that the variables are mutually independent or have Gaussian dependence structure, which is simple to model and to fit data. Additionally, some advanced UQ methodologies rely on a transformation to map the correlated input variables to variables with independent components. The Gaussian assumption and the associated correlation gives a convenient representation of input dependencies, but it may introduce a bias in the response estimates when the real dependence structure deviates from this assumption.

Dependence modeling has recently received widespread attention and adoptions in mathematics and engineering [62–68]. This is mainly due to the significant development of copula models and vine copulas in particular. Copula theory is used to separately model the dependence and the marginal distribution, but it is often limited to low-dimensional problems, typically bivariate or simple copula families. To overcome this limitation, Bedford and Cooke [69]

and Joe [70] first proposed the vine copulas to extend bivariate copulas to high dimensional problems. Consequently, this work makes use of the copula theory to construct accurate and objective probabilistic models for quantification of uncertainties resulting from the small datasets. Then these uncertainties with copula dependence modeling are efficiently propagated through the proposed methodology.

1.1.5 Imprecise global sensitivity analysis

Sensitivity analysis studies the relative impact of uncertain input parameters on uncertainty in the response of a system such that the relative importance of each stochastic input can be ranked. Reviews on various sensitivity methods can be found in [71, 72]. Generally, there are two classes of sensitivity analysis including *local sensitivity analysis (LSA)* and *global sensitivity analysis (GSA)*. LSA focuses on the influence of small variations of the input parameters around a certain value on the Quantify of Interest (QoI). GSA examines the overall influence of variations in the input parameters on system response. In other words, GSA provides a more comprehensive consideration of uncertainty associated with the model inputs. A large number of GSA studies and various engineering applications can be found in [73–78].

Relevant studies of *imprecise sensitivity analysis* are relatively scarce. Oberguggenberger et al. [79,80] discussed various classical and imprecise approaches

including fuzzy sets, random sets and p-boxes for sensitivity analysis in engineering. Helton et al. [81] provided a survey of available methods for uncertainty quantification (UQ) and sensitivity analysis. Helton et al. [72] also described several approaches for sensitivity analysis in the evidence theory based on the belief and plausibility measures. Li and Mahadevan [82, 83] described the GSA using Sobol indices in the presence of input uncertainty and model uncertainty. Schob and Sudret [84, 85] employed the P-boxes with polynomial chaos expansions for GSA in the context of imprecise probabilities. Using the proposed novel UQ methodology, this work aims at investigating the GSA associated with imprecise probability resulting from small datasets.

1.2 Organization of the Dissertation

In this dissertation, a novel UQ methodology is developed to address the challenge involving uncertainty quantification and propagation given small datasets. Chapter 2 shows the fundamental framework of the proposed methodology. Information-theoretic multimodel and Bayesian inference are employed to quantify both model-form and model parameter uncertainties. An optimal sampling density is derived to efficiently propagate the uncertainties identified from model selection and parameter estimation. A plate buckling strength problem illustrates the efficiency and adaptivity of the methodology, particu-

larly as additional data are collected. Chapter 3 provides an investigation into the influence of noninformative and informative prior probabilities on the resulting uncertainties. To overcome the limitation in assumption of variable correlation, a generalized study based on the proposed UQ methodology is performed to explore the uncertainty quantification and propagation of imprecise probabilities with copula dependence modeling, shown in Chapter 4. As one of the important research areas in UQ, sensitivity analysis plays a critical role in determining the influence of each of random input variables. Chapter 5 applies the developed UQ methodology to study the imprecise global sensitivity analysis (GSA) when data is scarce. The new imprecise GSA is implemented for assessing the influence of constituent material properties on overall composite material properties. Finally, a brief conclusion and future work are given by Chapter 6 .

Chapter 2

Quantification and efficient propagation of imprecise probabilities

This chapter addresses the problem of uncertainty quantification and propagation when data for characterizing probability models are very limited. A novel methodology is proposed to fully quantify the uncertainties associated with probability model-form and model parameters and then efficiently propagate these uncertainties. We introduce both information-theoretic multimodel inference, and Bayesian multimodel inference methods to identify the candidate probability models along with their model probabilities and employ Bayesian inference to estimate the joint posterior parameter densities for each plausi-

ble probability model. The full set of probability models are then propagated through an optimal importance sampling density that is representative of all plausible models, propagating this density and reweighting the samples based on each of the candidate models. The proposed methodology significantly reduces the computational cost compared with the conventional multiple-loop Monte Carlo simulation methods, as presented in Fig. 1.1.

2.1 Multimodel selection from small data

In this section, we review the principles of information-theoretic and Bayesian multimodel selection to assess the viability of various probability models to represent a dataset. Probability model selection is overcome in the information-theoretic method using the Akaike Information Criterion [56] to evaluate the viability of each model and estimate the probability that the given model is the "best" model. In the Bayesian method, a generalized Bayes' factor approach is presented for determining model probabilities.

2.1.1 Information-theoretic multimodel inference

Traditionally, statistical inference is applied to select a single "best" model given a set of candidate models and available data, and the model is the sole model used for making inference from data. Any uncertainty associated with

model selection is ignored because a single best model has been identified. However, such selection approaches often require very large datasets. In the case of small data, it is not straightforward (and often impossible) to identify a unique best model. Consequently, it is necessary to take into account the model uncertainty and compare the validity of multiple candidate models - a process referred to as multimodel inference [58]. We herein avoid the use of the term "true" model since we adhere to the information-theoretic belief that models are approximations of reality and hence a "true" model does not exist. On the contrary, the best model is the one that minimizes the difference between the model and observed reality. In this study, we employ a widely used approach developed by Burnham and Anderson [58] to quantify the model selection uncertainty by multimodel inference, as summarized below.

Generally speaking, model selection requires a well-justified criterion for selecting the best model. There are two popular approaches including the information-theoretic selection criteria utilizing Kullback-Leibler (K-L) information [86] and Bayesian model selection based on the Bayes' factor.

In the Bayesian setting, Bayes' factor B_F is defined as the ratio of the likelihoods of the data for the two models, M_0 and M_1:

$$\frac{p(M_0|\boldsymbol{d})}{p(M_1|\boldsymbol{d})} = \frac{p(M_0)p(\boldsymbol{d}|M_0)}{p(M_1)p(\boldsymbol{d}|M_1)} = B_F \cdot \frac{p(M_0)}{p(M_1)} \tag{2.1}$$

where

$$p(\boldsymbol{d}|M) = \int p(\boldsymbol{\theta}|M)p(\boldsymbol{d}|\boldsymbol{\theta}, M)d\boldsymbol{\theta} \tag{2.2}$$

It is noted that the above equation is generally interpreted that $B_F > 1$ implies that the data provides greater evidence in support of model M_0 and vice versa for $B_F < 1$. Bayes' factors are sensitive to the prior probabilities as shown in Eq. (2.1). Particularly in the limited data case, infinite Bayes factors exist for a given dataset based on the assignment of the priors - only some of which are reasonable. Another well-known related metric for Bayesian model selection is the Bayesian information Criterion (BIC), which is derived from the integrated likelihood function. The BIC is derived by estimating $-2\log(p(\boldsymbol{d}|M))$ through a Taylor series expansion around the maximum likelihood estimate of the parameters, $\hat{\boldsymbol{\theta}}$, with some terms neglected as [25]:

$$\text{BIC} = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + K\log n \tag{2.3}$$

where K is the dimension of the parameter vector $\boldsymbol{\theta}$, n is the sample size of the dataset and $\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M) = p(\boldsymbol{d}|\hat{\boldsymbol{\theta}}, M)$ is the likelihood function given the maximum likelihood estimate of the parameters $\hat{\boldsymbol{\theta}}$. The BIC also provides a convenient means of estimating a Bayes' factor for model comparison with some implicitly defined prior distribution that may or may not be reasonable [87].

To achieve the model selection using BIC, it is necessary to construct a rel-

ative scale:

$$\Delta_B^{(i)} = \text{BIC}^{(i)} - \text{BIC}^{\min} \tag{2.4}$$

where $\text{BIC}^{(i)}$ is the BIC for candidate model M_i and $\text{BIC}^{\min} = \min_i(\text{BIC}^{(i)})$. This normalizes the best model to a value $\Delta_B^{(i)} = 0$. For such case, we can establish the posterior model probabilities, $\mathfrak{p}_i$, as [25]:

$$\mathfrak{p}_i = p(M_i|\boldsymbol{d}) = \frac{\exp(-\frac{1}{2}\Delta_B^{(i)})}{\sum_{i=1}^{N}\exp(-\frac{1}{2}\Delta_B^{(i)})} \tag{2.5}$$

under the assumption that the prior models M_i, have equal probability $\frac{1}{N}$. As noted by Burnham and Anderson [58], it does not imply that a true model is among the candidate models M_i or that such a model even exists - even if $\mathfrak{p}_i = 1$. Instead, these probabilities can be interpreted as the probability that M_i is the model that the BIC would select with $\mathfrak{p}_i = 1$ given $n \to \infty$ (referred to as the BIC target model).

In terms of the information-theoretic methodology, model selection is implemented by establishing a criterion for the information loss resulting from approximating truth with a model. As a result, an appropriate model selection criterion is to minimize the information loss. Based on this idea, Akaike proposed the Akaike Information Criterion (AIC) based on the fact that the expected relative K-L information could be approximated by the maximized

log-likelihood function with a bias correction [56]. AIC is defined as follows:

$$\text{AIC} = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + 2K \tag{2.6}$$

where $2K$ is a bias correction factor. As an information-theoretic criterion, AIC provides a simple estimate of the K-L information, but it can also be conceived in a Bayesian context by the use of a class of *savvy priors* (more details can be found in [58]). Similar to the BIC, it is also necessary to establish a relative scale for AIC values, as described below:

$$\Delta_A^{(i)} = \text{AIC}^{(i)} - \text{AIC}^{\min} \tag{2.7}$$

Consequently, the best model has $\Delta_A^{(i)} = 0$ but all other models have positive $\Delta_A^{(i)}$ interpreted as the information lost relative to the best model. Furthermore, the simple transformation $\exp(-\frac{\Delta_A^{(i)}}{2})$ provides the likelihood for the model M_i given the data. We can therefore estimate the model probability by normalizing these likelihoods as:

$$\pi_i = p(M_i|d) = \frac{\exp(-\frac{1}{2}\Delta_A^{(i)})}{\sum_{i=1}^{N}\exp(-\frac{1}{2}\Delta_A^{(i)})} \tag{2.8}$$

In fact, these probabilities can be interpreted as the probability that model M_i is the K-L best model for the data.

Both the BIC and AIC are asymptotic estimators that require large datasets - each with their own distinct advantages and disadvantages [58, 87]. To handle small datasets, a critical extension of the AIC has been proposed [88, 89] , named as, AIC_c, which introduces a second-order bias correction term yielding:

$$\text{AIC}_c = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + 2K + \frac{2K(K+1)}{n-K-1}. \tag{2.9}$$

AIC_c is often used if $\frac{n}{K} <\sim 40$. Since $\text{AIC}_c \to \text{AIC}$ as $n \to \infty$, it usually makes sense to adopt the second-order correction version.

In this study, we make use of the concept of AIC_c for multimodel inference due to its Bayesian interpretation, information-theoretic property and allowance for small data size. Probability models are ranked in terms of $\Delta_A^{(i)}$ and the corresponding model probabilities are assigned based on Eq. (2.8).

2.1.2 Bayesian multimodel inference

Probabilistic model selection has been widely attractive with the development of methods based on information theory and Bayes' theory. The information-theoretic approach, derived from the work of Akaike [56, 90] and its further generalization [88,91,92], has been discussed in the previous section. This section will focus on the Bayesian approaches which involve the notion of posterior model probabilities proposed by Raftery [93] and recently have been revived by

Beck [94–96] and Oden [97–99].

2.1.2.1 Generalization of Bayes' factor

As mentioned in the previous section, Bayes' factor is often employed for selection between two models M_i and M_j given data $\boldsymbol{d}$ through the expression

$$\underbrace{\frac{p(M_i|\boldsymbol{d})}{p(M_j|\boldsymbol{d})}}_{\text{Posterior odds}} = \underbrace{\frac{p(M_i)}{p(M_j)}}_{\text{Prior odds}} \times \underbrace{\frac{p(\boldsymbol{d}|M_i)}{p(\boldsymbol{d}|M_j)}}_{\text{Bayes' factor}} \tag{2.10}$$

where, again, Bayes' factor is defined as the ratio of the evidence of M_i and M_j, and the prior odds is the ratio of model prior of M_i and M_j. If the posterior odds are greater than one, then model M_i is selected while if the posterior odds are less than one, model M_j is selected.

Intuitively, it is not difficult to generalize the Bayes' factor concept for comparison of multiple candidate models. Consider the aforementioned collection of m parametric models $\boldsymbol{\mathcal{M}}$, with each model M_j having an associated prior probability $\lambda_j = p(M_j)$ with $\sum_{j=1}^{m} \lambda_j = 1$. The posterior model probabilities π_j can be calculated based on the prior model probabilities λ_j via the following formulation

$$\pi_j = p(M_j|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|M_j)p(M_j)}{\sum_{k=1}^{m} p(\boldsymbol{d}|M_k)p(M_k)}, \quad j = 1, \ldots, m \tag{2.11}$$

having $\sum_{j=1}^{m} \pi_j = 1$ and where

$$p(\boldsymbol{d}|M_j) = \int_{\Theta_j} p(\boldsymbol{d}|\boldsymbol{\theta}_j, M_j) p(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j, \quad j = 1, \ldots, m \tag{2.12}$$

where $p(\boldsymbol{d}|M_j)$ is referred as to the evidence of model M_j.

Typically, the model $M_k \in \boldsymbol{\mathcal{M}}$ with highest probability $p(M_k|\boldsymbol{d})$ is identified as the most plausible in the set $\boldsymbol{\mathcal{M}}$ for the given data $\boldsymbol{d}$. Instead of selecting the model with the highest probability, multimodel inference aims to rank the models according to their posterior model probabilities given by Eq. (2.11) and retain all the plausible models with non-negligible probability.

In fact, model parameter estimation using Bayesian inference does not require the evidence $p(\boldsymbol{d}|M_j)$ to be computed using the MCMC algorithm. However, the evidence $p(\boldsymbol{d}|M_j)$ is critical in Bayesian multimodel inference, as evident from Eq. (2.12), and consequently needs to be calculated with caution. We discuss how to calculate the evidence in the following section.

2.1.2.2 Bayesian evidence calculation

There are several different approaches for the calculation of evidence in Eq. (2.11). The integral in Eq. (2.11) is rarely evaluated analytically and the most common way is to use the approximate or statistically exact (i.e. Monte Carlo) approaches instead.

One efficient approximation method is Laplace's approach [92] that approximates the evidence $p(\boldsymbol{d}|M_j)$ using the following formulation

$$p(\boldsymbol{d}|M_j) \approx \exp\left\{\log(p(\boldsymbol{d}|\boldsymbol{\theta}_j^*, M_j))\right\} p(\boldsymbol{\theta}_j^*|M_j)(2\pi)^{K_j/2} n^{-K_j/2} |H^*(\boldsymbol{\theta}_j^*)|^{-1/2} \tag{2.13}$$

Taking the logarithm of this expression and multiplying it by -2, we obtain

$$\begin{aligned} -2\log(p(\boldsymbol{d}|M_j)) \approx & -2\log(p(\boldsymbol{d}|\boldsymbol{\theta}_j^*, M_j)) + K_j \log(n) + \log|H^*(\boldsymbol{\theta}_j^*)| - K_j \log(2\pi) \\ & - 2\log(p(\boldsymbol{\theta}_j^*|M_j)) \end{aligned} \tag{2.14}$$

where $\boldsymbol{\theta}_j^*$ is the maximum likelihood estimate, H^* is the inverse Hessian of the negative log likelihood (Fisher information matrix) and K_j is the dimension of the parameter vector $\boldsymbol{\theta}$. Ignoring the terms in Eq. (2.14) with order less than $O(1)$ for large sample size n yields the Bayesian Information Criteria (BIC) [91] as shown in Eq. (2.3). This quantity can be used to construct an asymptotic approximation to Bayes' factor, namely $\text{BF}_{i,j} \approx \exp(-(\text{BIC}_i - \text{BIC}_j))/2$ [93]. Integrated with the model prior $\lambda_j = p(M_j)$, posterior model probabilities from Eq. (2.11) can be given by

$$\pi_j^{\text{BIC}} \approx \frac{\exp(-\frac{1}{2}(\text{BIC}_j - \text{BIC}_{\min}))\lambda_j}{\sum_{k=1}^{m} \exp(-\frac{1}{2}(\text{BIC}_k - \text{BIC}_{\min}))\lambda_k} \tag{2.15}$$

where $\text{BIC}_{\min} = \min_j(\text{BIC}_j)$. Assigning uniform prior model probabilities to the

set $\boldsymbol{\mathcal{M}}$, $\lambda_j \equiv 1/m$, yields what are referred to as BIC model weights. Eq. (2.15) can actually be deemed as generalized BIC model weights for arbitrary prior model probabilities. It is also noted that Eq. (2.3) may be thought of as an implicit approximation to evidence $p(\boldsymbol{d}|M_j)$ given a *noninformative parameter prior* (or *Jeffreys parameter prior*) even though it does not explicitly depend on a parameter prior.

Multimodel inference using the information-theoretic method, introduced in [58] and introduced in the previous section, can be thought of as a special case of the Bayesian evidence-based multimodel selection. Akaike Information Criterion (AIC) has been discussed previously and the model probabilities using AIC are defined as shown in Eq. (2.8). As shown by [58], Eq. (2.8) is in fact a special case of Eq. (2.15) in which the prior model probabilities λ_j take the form

$$\lambda_j = \frac{\exp\left(\frac{1}{2}K_j \log(n) - K_j\right)}{\sum_{k=1}^{m} \exp\left(\frac{1}{2}K_k \log(n) - K_k\right)} \tag{2.16}$$

This form of priors are referred to as *savvy* (shrewdly informed) priors because they depend on n and K_k.

Model probabilities using AIC and BIC are crucial since they show directly the impact of priors in the asymptotic case. While the model probabilities in Eq. (2.11) are general, they can be approximated (in large data cases) by Eq. (2.15) and the AIC derived model probabilities are an instance of this approximation under certain prior information.

As asymptotic quantities, both AIC and BIC require large dataset size such that their application is often limited in practice, particularly in the small data case considered here. Even though Hurvich et al. [88, 89] has proposed a correction of the AIC for small data case as discussed in the previous section, this again implies a certain prior form and one objective here is to study the effect of prior probability. As a result, we have to consider other estimators for Eq. (2.12), which do not need to assume a specified prior form or asymptotic conditions. In this work, we use a Monte Carlo-based statistical estimator

$$\hat{p}(\boldsymbol{d}|M_j) = \frac{1}{N_k}\sum_{k=1}^{N_k} p(\boldsymbol{d}|\boldsymbol{\theta}_j^k, M_j), \quad \boldsymbol{\theta}_j^k \sim p(\boldsymbol{\theta}_j|M_j), \quad j = 1, \ldots, m \tag{2.17}$$

where N_k is the number of samples. The samples $\boldsymbol{\theta}_j^k$ are drawn from the parameter prior distribution. The Monte Carlo algorithm for evidence calculation used in this work has an acceptable computational cost. The computational efficiency can be further improved with parallel and high performance computing. If high dimensional or complex models are considered, MCMC-based algorithms, i.e. nesting sampling [100] as well as Chib and Jeliazkov [101] may be a better choice as suggested in the recent literature review [102–104].

2.2 Bayesian inference and parametric uncertainty

Once the model-form uncertainties are identified and quantified, then we need to focus on the uncertainties associated with the model parameters. Consider the random variable $\boldsymbol{X}$ with probability model M and uncertain parameters $\boldsymbol{\theta}$. Bayes' rule assigns a prior probability density function (pdf), $p(\boldsymbol{\theta}; M)$ for the model parameters $\boldsymbol{\theta}$. The prior pdf reflects the existing knowledge or belief about the parameter distribution. Given collected data, Bayes' rule updates our knowledge of the parameters $\boldsymbol{\theta}$ for M to give a posterior distribution $p^*(\boldsymbol{\theta}|\boldsymbol{d}, M)$ reflecting our updated knowledge by:

$$p^*(\boldsymbol{\theta}|\boldsymbol{d}, M) = \frac{p(\boldsymbol{d}|\boldsymbol{\theta}, M)p(\boldsymbol{\theta}; M)}{p(\boldsymbol{d}; M)} \propto \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d}, M)p(\boldsymbol{\theta}; M) \tag{2.18}$$

where $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d}, M) = p(\boldsymbol{d}|\boldsymbol{\theta}, M)$ is referred to as the likelihood function and normalizing factor $p(\boldsymbol{d}; M)$, means the evidence, which is calculated by marginalizing $\mathcal{L}(\cdot)$ over the parameters $\boldsymbol{\theta}$

$$p(\boldsymbol{d}; M) = \int \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d}, M)p(\boldsymbol{\theta}; M)d\boldsymbol{\theta} \tag{2.19}$$

This is not a trivial task as the integral in Eq. (2.19) is usually analytically intractable. Certain special cases may have closed-form solutions. These cases are called conjugate distributions and the prior and posterior distribution belong to the same special family. Most conjugate relations are from the exponential distribution families and one can find more details in [105]. More generally, numerical methods are employed to solve this intractable, multidimensional integral. As one of most popular approaches, Markov Chain Monte Carlo (MCMC) is often used to compute the evidence and generate posterior samples from $p^*(\boldsymbol{\theta}|\boldsymbol{d}, M)$. The most common used MCMC algorithms are Gibbs sampling [106] and Metropolis-Hastings (MH) [107]. In this work, we utilize the affine invariant ensemble sampler for MCMC proposed by Goodman and Weare [108, 109]. More details about this algorithm are shown in Appendix A.

In the context of small data, Bayesian inference improves the parameter estimates over prior information only to a limited extent. For such cases, the posterior parameter estimation will likely possess large variance that does not instill confidence in the selection of a single point estimator, i.e. maximum a posterior (MAP) estimator or maximum likelihood estimate. Instead, in this work, we propose to retain the complete joint parameter density and then propagate it through the model for response output.

2.3 Optimal importance sampling for multimodel uncertainty propagation

In this section, we turn our attention to the study of efficient uncertainty propagation, particularly given the case of small datasets. A large number of related methods for uncertainty propagation have been proposed but we are presently unaware of any approaches capable of simultaneously propagating many probability models without explicitly propagating each distribution individually at large computational expense [50]. In this work, we aim at utilizing the concepts of importance sampling to achieve the simultaneous uncertainty propagation of the multiple possible models identified through multimodel selection with Bayesian inference.

2.3.1 Importance sampling

Let us define a generic performance function $g(\boldsymbol{X})$ as the response quantity for a specific system. Uncertainty propagation is generally concerned with evaluating the expected value $E[g(\boldsymbol{X})]$ where $\boldsymbol{X} \in \Omega$ is a random vector having probability model $M_{\boldsymbol{X}}$ with density function $p(\boldsymbol{x})$. Often, Monte Carlo analysis

is used to estimate

$$E_p[g(\boldsymbol{X})] = \int_\Omega g(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i) \tag{2.20}$$

where N is the number of samples, $\boldsymbol{x}_i$ are independent random samples drawn from $p(\boldsymbol{x})$, and $E_p[\cdot]$ is the expectation with respect to $p(\boldsymbol{x})$. It sometimes may not be easy to generate samples directly from $p(\boldsymbol{x})$, and for such case we may consider an alternate density $q(\boldsymbol{x})$ which is easier to sample. That is the original principle of importance sampling. The Monte Carlo estimator in Eq. (2.20) is then modified to correct the bias generated by sampling from an alternate distribution as:

$$E_q\left[g(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] = \int_\Omega g(\boldsymbol{x})\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}q(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i)\frac{p(\boldsymbol{x}_i)}{q(\boldsymbol{x}_i)} \tag{2.21}$$

where $E_q[\cdot]$ is the expectation with respect to $q(\boldsymbol{x})$. As the importance weights, the ratios $w(\boldsymbol{x}_i) = p(\boldsymbol{x}_i)/q(\boldsymbol{x}_i)$ play a critical role in the proposed methodology of this work.

2.3.2 Optimal sampling density for a single target density

Many studies have been proposed toward identifying an efficient proposal sampling density $q(\boldsymbol{x})$ given a known target density $p(\boldsymbol{x})$. Most cases focus on the goal of variance reduction. A classical optimal sampling density is given by $q(\boldsymbol{x}) = \frac{g(\boldsymbol{x})p(\boldsymbol{x})}{E_p[g(\boldsymbol{x})]}$ which achieves a zero variance estimator but is always infeasible in practice.

Instead of achieving a variance reduction, we are more interested in ensuring that our sampling density is as close as possible to the target density $p(\boldsymbol{x})$, given the difficulty of sampling from $p(\boldsymbol{x})$ itself. This is achieved by minimizing the f-divergence [110–112] which defines the difference between two distributions P and Q over a space Ω with measure μ as:

$$D_f(P \parallel Q) = \int_\Omega f\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) q(\boldsymbol{x}) d\mu(\boldsymbol{x}) \tag{2.22}$$

Various functions $f(\cdot)$ have been proposed based on the basic definition in Eq. (2.22), for example, Kullback-Leibler divergence [86], the Hellinger distance [113], and the total variation distance [114]. In this work, the Hellinger

distance is firstly employed

$$H^2(P \parallel Q) = \frac{1}{2}\int_{\Omega}\left(\sqrt{p(\boldsymbol{x})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x}. \tag{2.23}$$

The objective then is to identify a form for $q^*(\boldsymbol{x})$ that minimizes this difference for a given family of distributions. This yields the square Minimum Hellinger Distance Estimator (MHDE) [115,116] relative to the target density $p(\boldsymbol{x})$ given by:

$$q^*(\boldsymbol{x}) = \arg\min \frac{1}{2}\int_{\Omega}\left(\sqrt{p(\boldsymbol{x})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x} \tag{2.24}$$

2.3.3 Optimal sampling density for multiple distributions

In terms of the case of imprecise probabilities here, the target density is not uniquely defined (i.e. it is not known precisely). Instead, multiple probability models, M_i, are plausible, each with uncertain (probabilistic) parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, quantified through Bayesian inference and density functions $p_i(\boldsymbol{x}|\boldsymbol{\theta})$ having model probabilities π_i identified through multimodel inference (Eq. (2.8) or Eq. (2.11)). Our ultimate object is to identify a *single* proposal sampling density $q^*(\boldsymbol{x})$ that is well representative of all target densities $p_i(\boldsymbol{x}|\boldsymbol{\theta})$ such that we can use $q^*(\boldsymbol{x})$ with importance sampling to simultaneously propagate the

full set of plausible candidate probability models. Assume that we have a finite set $\mathbb{M} = \{M_i\}, i = 1, 2, ..., N_d$ of candidate target probability models having densities $p_i(\boldsymbol{x}|\boldsymbol{\theta})$. The total Hellinger distance can therefore be formulated as:

$$\hat{H}^2(\mathbb{M} \parallel Q) = \sum_{i=1}^{N_d} H^2(M_i \parallel Q) = \sum_{i=1}^{N_d} \frac{1}{2} \int_{\Omega} \left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})} \right)^2 d\boldsymbol{x} \tag{2.25}$$

The total Hellinger distance is thus a random variable indexed on $\boldsymbol{\theta}$ with expected value presented by:

$$\begin{aligned} E\left[\hat{H}^2(\mathbb{M} \parallel Q)\right] = \sum_{i=1}^{N_d} E\left[H^2(M_i \parallel Q)\right] &= \sum_{i=1}^{N_d} \frac{1}{2} \int_{\Omega} E_{\theta}\left[\left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})} \right)^2 \right] d\boldsymbol{x} \\ &= E_{\theta}\left[\int_{\Omega} \sum_{i=1}^{N_d} \frac{1}{2} \left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})} \right)^2 d\boldsymbol{x} \right] \end{aligned} \tag{2.26}$$

Eq. (2.26) is referred to as the total expected squared Hellinger distance.

To find an optimal sampling density $q(\boldsymbol{x})$, we define an overall optimization problem that is to minimize the total expected squared Hellinger distance expressed as a functional $\mathcal{T}(q)$ given the isoperimetric constraint $\mathcal{I}(q)$ as

$$\begin{aligned} &\underset{q}{\text{minimize}} \quad \mathcal{T}(q) = E_{\theta}\left[\int_{\Omega} F(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) d\boldsymbol{x} \right] \\ &\text{subject to} \quad \mathcal{I}(q) = \int_{\Omega} q(\boldsymbol{x}) d\boldsymbol{x} - 1 = 0 \end{aligned} \tag{2.27}$$

where

$$F(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) = \frac{1}{2} \sum_{i=1}^{N_d} \left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})} \right)^2 \tag{2.28}$$

and the constraint guarantees that $q(\boldsymbol{x})$ is a valid probability density function. The optimization problem here allows us to apply the Lagrange multipliers method with the calculus of variations. We define the Lagrangian as:

$$\mathcal{L}(q, \lambda) = \mathcal{T}(q) + \lambda \mathcal{I}(q) \tag{2.29}$$

The function $q(\boldsymbol{x})$ that minimizes the functional $\mathcal{T}(q)$ leads to the variation of the action functional $\mathcal{L}(q, \lambda)$ in terms of q and λ to vanish. Hence, the variation of the action functional $\mathcal{L}(q, \lambda)$ is expressed as:

$$\begin{aligned}
\delta\mathcal{L}(q, \lambda) &= \delta\mathcal{T}(q) + \delta(\lambda\mathcal{I}(q)) \\
&= E_\theta \left[\int_\Omega \frac{\partial F}{\partial q}(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})\delta q(\boldsymbol{x}))d\boldsymbol{x} \right] + \lambda \left(\int_\Omega \delta q(\boldsymbol{x}) d\boldsymbol{x} \right) + \mathcal{I}(q)\delta\lambda \\
&= E_\theta \left[\int_\Omega \left(\frac{\partial F}{\partial q}(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) + \lambda \right) \delta q(\boldsymbol{x}) d\boldsymbol{x} \right] + \left(\int_\Omega q(\boldsymbol{x}) d\boldsymbol{x} - 1 \right) \delta\lambda
\end{aligned} \tag{2.30}$$

The formulation above has to vanish for all variations $\delta\lambda$ and δq according to the basic principle of the calculus of variations, that is equivalent to

$$\begin{aligned}
E_\theta \left[\frac{\partial F}{\partial q}(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) + \lambda \right] = 0, \\
\int_\Omega q(\boldsymbol{x}) d\boldsymbol{x} - 1 = 0
\end{aligned} \tag{2.31}$$

If the total expected squared Hellinger distance is considered, the expression

is given by

$$\begin{aligned}\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x})) &= \frac{1}{2}\frac{\partial\left(\sum_{i=1}^{N_d}\left(p_i(\boldsymbol{x}|\boldsymbol{\theta})-2\sqrt{(p_i(\boldsymbol{x}|\boldsymbol{\theta})q(\boldsymbol{x}))}+q(\boldsymbol{x})\right)\right)}{\partial q}\\ &= \frac{1}{2}\sum_{i=1}^{N_d}\left(-\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}+1\right)\\ &= \frac{N_d}{2}-\frac{1}{2}\sum_{i=1}^{N_d}\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}\end{aligned} \tag{2.32}$$

Considering the expectation with respect to $\boldsymbol{\theta}$ yields:

$$\begin{aligned}E_\theta\left[\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))+\lambda\right] &= E_\theta\left[\frac{N_d}{2}-\frac{1}{2}\sum_{i=1}^{N_d}\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}+\lambda\right]\\ &= \frac{N_d+2\lambda}{2}-\frac{1}{2}\sum_{i=1}^{N_d}E_\theta\left[\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}\right]\end{aligned} \tag{2.33}$$

The optimal sampling density with minimized expectation can be derived by setting Eq. (2.33) equal to zero

$$q^*(\boldsymbol{x}) = \left(\frac{1}{N_d+2\lambda}\right)^2\left(\sum_{i=1}^{N_d}E_\theta\left[\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}\right]\right)^2 \tag{2.34}$$

where λ is selected to ensure $\int_\Omega q^*(\boldsymbol{x})d\boldsymbol{x}=1$. Unfortunately, in this case, it is not straightforward to estimate λ given the multiple (possibly large) number of probability density distributions $p_i(\boldsymbol{x}|\boldsymbol{\theta})$.

Another important metric in f-divergence is the mean square difference

(MSD), that is given by

$$\mathcal{M}(P \parallel Q) = \frac{1}{2} \int_{\Omega} \left(p(\boldsymbol{x}) - q(\boldsymbol{x})\right)^2 d\boldsymbol{x} \tag{2.35}$$

The corresponding total expected mean squared difference can be expressed as

$$\begin{aligned} E\left[\mathcal{M}(\mathbb{M} \parallel Q)\right] = \sum_{i=1}^{N_d} E\left[\mathcal{M}(M_i \parallel Q)\right] &= \sum_{i=1}^{N_d} \frac{1}{2} \int_{\Omega} E_{\theta}\left[\left(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x})\right)^2\right] d\boldsymbol{x} \\ &= E_{\theta}\left[\sum_{i=1}^{N_d} \frac{1}{2} \int_{\Omega} \left(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x})\right)^2 d\boldsymbol{x}\right] \\ &= E_{\theta}\left[\int_{\Omega} \sum_{i=1}^{N_d} \frac{1}{2} \left(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x})\right)^2 d\boldsymbol{x}\right] \end{aligned} \tag{2.36}$$

Let us formulate the optimization with respect to the functional $\hat{\mathcal{T}}(q)$ and and the constraint $\hat{\mathcal{I}}(q)$ as follows:

$$\begin{aligned} &\underset{q}{\text{minimize}} \quad \hat{\mathcal{T}}(q) = E_{\theta}\left[\int_{\Omega} \hat{F}(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) d\boldsymbol{x}\right] \\ &\text{subject to} \quad \hat{\mathcal{I}}(q) = \int_{\Omega} q(\boldsymbol{x}) d\boldsymbol{x} - 1 = 0 \end{aligned} \tag{2.37}$$

where the action functional $\hat{F}$ is:

$$\hat{F}(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) = \frac{1}{2} \sum_{i=1}^{N_d} \left(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x})\right)^2 \tag{2.38}$$

Similarly, we define the Lagrangian associated with the objective functional

$\hat{\mathcal{T}}(q)$ and the constraint $\hat{\mathcal{I}}(q)$, then eliminate all variations δq, $\delta\lambda$

$$\begin{aligned}\delta\hat{\mathcal{L}}(q,\lambda) &= \delta\hat{\mathcal{T}}(q) + \delta(\lambda\hat{\mathcal{I}}(q)) \\ &= E_\theta\left[\int_\Omega\left(\frac{\partial\hat{F}}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))+\lambda\right)d\boldsymbol{x}\right]\delta q + \left(\int_\Omega q(\boldsymbol{x})d\boldsymbol{x}-1\right)\delta\lambda\end{aligned} \tag{2.39}$$

which leads to $\int_\Omega q(\boldsymbol{x})d\boldsymbol{x}=1$ and

$$\begin{aligned}E_\theta\left[\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))+\lambda\right] &= E_\theta\left[-\sum_{i=1}^{N_d}\left(p_i(\boldsymbol{x}|\boldsymbol{\theta})-q(\boldsymbol{x})\right)+\lambda\right] \\ &= q(\boldsymbol{x})N_d+\lambda-\sum_{i=1}^{N_d}E_\theta\left[p_i(\boldsymbol{x}|\boldsymbol{\theta})\right] \\ &= 0\end{aligned} \tag{2.40}$$

and solving for $q(\boldsymbol{x})$ gives the minimizer is

$$q^*(\boldsymbol{x}) = \frac{1}{N_d}\left(\sum_{i=1}^{N_d}E_\theta\left[p_i(\boldsymbol{x}|\boldsymbol{\theta})\right]-\lambda\right) \tag{2.41}$$

which leads to $\lambda = 0$ such that $\int_\Omega q^*(\boldsymbol{x})d\boldsymbol{x} = 1$. The final optimal sampling density is therefore,

$$q^*(\boldsymbol{x}) = \frac{1}{N_d}\sum_{i=1}^{N_d}E_\theta\left[p_i(\boldsymbol{x}|\boldsymbol{\theta})\right] \tag{2.42}$$

The solution in Eq. (2.42) is a mixture distribution combining the candidate target densities and their parameter ranges. It assumes that each probability model M_i is equally probable (i.e. $\pi_i = \frac{1}{N_d}$). It is straightforward to show that

this solution generalizes as:

$$q^*(\boldsymbol{x}) = \frac{1}{N_d} \sum_{i=1}^{N_d} E_\theta \left[\pi_i \cdot p_i(\boldsymbol{x}|\boldsymbol{\theta}) \right] \tag{2.43}$$

where π_i is the AIC_c or Bayesian model probability (see Eq. (2.8) or Eq. (2.11)) for model M_i satisfying $\sum_{i=1}^{N_d} \pi_i = 1$.

2.4 Proposed methodology for quantification and propagation of imprecise probabilities

We herein summarize the developed methodology outlined in the previous chapters and provide a flowchart as shown in Fig. 2.1.

- *Step 1: Multimodel inference* - Given a small dataset $\boldsymbol{d}$, identify the set of candidate probability models $\mathbb{M} = \{M_i\}, i = 1, ..., N_d$. Apply information-theoretic or Bayesian multimodel inference to compute the corresponding model probabilities π_i for each candidate model M_i, based on Eqs. (2.7) - (2.9) or Eq. (2.11). Plausible models with extremely low probability might be discarded.

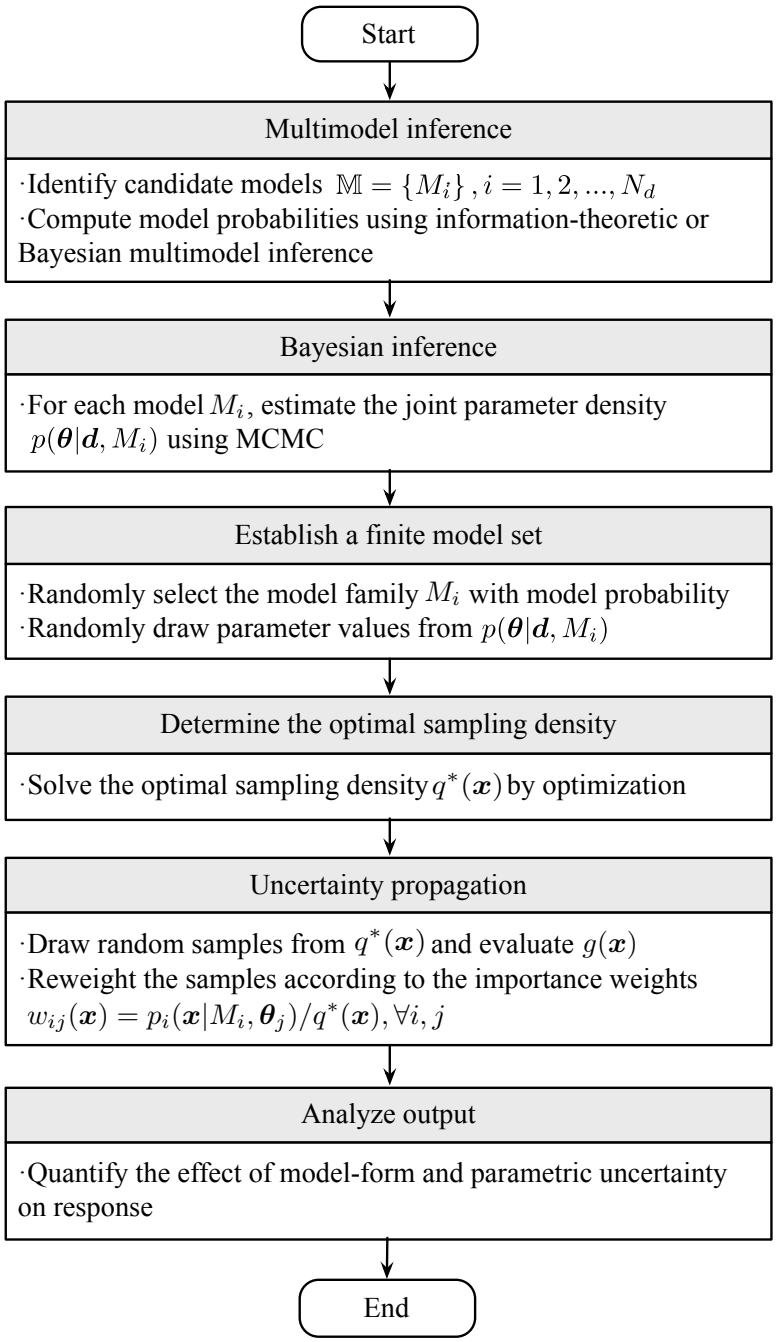


Figure 2.1: Flowchart of the proposed method for propagation of imprecise probabilities.

- *Step 2: Bayesian inference* - For each candidate probability model M_i, employ Bayes' rule to estimate the joint pdf of the model parameters, $p(\boldsymbol{\theta}|\boldsymbol{d}, M_i), i = 1, ..., N_d$. Generally, this is achieved by the MCMC algortihm.

- *Step 3: Build a finite model set* - Theoretically, Steps 1 - 2 yield an infinite set of parametrized probability models. This set can be reduced to a finite set of models using a conventional Monte Carlo sampling method. Regarding each model in the finite set, the model family M_i is chosen randomly with probability π_i. The model parameter values are then randomly drawn from $p(\boldsymbol{\theta}|\boldsymbol{d}, M_i)$. Note that it may be advantageous to sample directly from the previous MCMC draws implemented in Step 2. It is critical to generate an adequate large model set to span the full range of the candidate models. No additional model evaluations are needed in order to consider additional densities.

- *Step 4: Determine the optimal sampling density* - The importance sampling density is determined through formulation and optimization as described in Section 2.3.3. This work utilizes the analytical solution of optimal density $q^*(\boldsymbol{x})$ from Eq. (2.42) derived by minimizing the expected mean squared differences, which comes at little computational cost.

- *Step 5: Uncertainty propagation* - We propagate the uncertainty using importance sampling reweighting algorithm. Samples are drawn from $q^*(\boldsymbol{x})$ and are re-weighted based on the importance weights $w_{ij}(\boldsymbol{x}) = \frac{p_i(\boldsymbol{x}|M_i,\boldsymbol{\theta}_j)}{q^*(\boldsymbol{x})}$. In other words, each sample drawn from $q^*(\boldsymbol{x})$ is re-weighted a large number of times according to each plausible probability model. Using these reweighted sets, statistical analysis is implemented across all candidate models to analyze the response statistics, including mean, standard deviation, distributions and probability failure, etc. corresponding to each candidate model.

2.5 Model updating

The developed method here provides a high degree of flexibility and consequently it is easily and adaptively updated to accommodate additional new collected data or new candidate probability models. The only potential limitation in this updating is a loss of optimality in the importance sampling density.

2.5.1 Adding data

Often, the original dataset $\boldsymbol{d}$ are so small that they produce large uncertainties in the output response. Additional data $\boldsymbol{d}^*$ can be gradually collected using the same testing methods in order to reduce these uncertainties. The

multimodel inference proposed in this work can be directly updated by computing the new model probabilities and the corresponding model parameter joint density can be simply updated by Bayesian updating. As a result, this will possibly increase the confidence in the model-form selection while also narrowing the posterior joint parameter densities. But, this updating comes at the expense of optimality in the importance sampling density. In other words, this might lead to an increase in the variance of statistical estimates from the importance sampling reweighting algorithm. This is also because the updating simply reweights the samples without recomputing the optimal sampling density. But in some cases, the variance increase may warrant additional samples, so a new optimal sampling density may be reconstructed. Additional samples drawn from this updated density in this case. The aggregate sample set (new and old samples) can be combined using mixture resampling algorithm [117] or multiple importance sampling method [118]. This will not be discussed further here.

2.5.2 Adding probability models

If the output confidence range becomes artificially narrow, it is potentially caused by undersampling the candidate probability model space. As mentioned previously, additional probability models can be introduced from the model-form and model parameter probabilities without necessarily requiring a new

optimal sampling density. If a new optimal sampling density is necessary, additional samples may be generated as discussed in the above section.

2.5.3 Notable limitations

Even with the flexibility discussed herein, the proposed methodology still has its limitations. The most notable one is that the set of candidate probability models has been established using only well-known distributions with a parametric form. However, nonparametric models, which have more general forms, are not currently considered and pose a few challenges, i.e. how to accurately identify a non-parametric model from very limited data without overfitting. Furthermore, the method is established based on the candidate probability models and the users can supply any necessary and appropriate model that they consider. When sufficiently diverse candidate probability models are available, the set of these models may be sufficient to span the epistemic uncertainty. But, when the user does not supply a sufficient set of candidate models, the method will likely underestimate or incorrectly predict uncertainties. As discussed in Section 2.5.1 and 2.5.2, the proposed methodology is robust enough that additional new candidate probability model can be added a posteriori at almost no additional cost.

2.6 Application to plate buckling strength problem

Uncertainty in the geometric and material properties of structural components can have a significant impact on the reliability and safety of the structural system. This section mainly illustrates the proposed methodology through uncertainty quantification and propagation in buckling strength of a simply supported plate under uniaxial compression. An analytical formulation for the normalized buckling strength was first proposed by Faulkner [119].

$$\psi = \frac{\sigma_u}{\sigma_0} = \frac{2}{\lambda} - \frac{1}{\lambda^2} \tag{2.44}$$

where σ_u is the ultimate stress at failure, σ_0 is the yield stress, and λ is the slenderness of the plate with width b, thickness t, and elastic modulus E given by

$$\lambda = \frac{b}{t}\sqrt{\frac{\sigma_0}{E}} \tag{2.45}$$

Eq. (2.45) was modified by Carlsen [120] to investigate the effect of residual stresses and non-dimensional initial deflection δ_0 associated with welding

$$\psi = \left(\frac{2.1}{\lambda} - \frac{0.9}{\lambda^2}\right)\left(1 - \frac{0.75\delta_0}{\lambda}\right)\left(1 - \frac{2\eta t}{b}\right) \tag{2.46}$$

where ηt is the width of the zone of tension residual stress.

The buckling strength in Eq. (2.46) is calculated according to the nominal values of six variables describing geometric and material properties as shown in Table 2.1. Due to uncertainties in the material and geometric properties, the actual values of these variables are usually different from the nominal design values. Consequently, it is interesting to study the effect of the six uncertain variables presented in Table 2.1 on the buckling strength prediction. Even though the extension to the multiple dimensional cases (full 6-variable in this example) is straightforward, we focus on the study of a single parameter case for brevity in illustration. Using global sensitivity analysis, we identify that the yield strength σ_0 is the most influential variable on the buckling strength among the six variables such that we consider only the influence of uncertainty in this critical material parameter.

Table 2.1: Statistical properties of plate material, geometry and imperfection variables from Hess [1] and Guedes Soares [2]

Variables	Physical Meaning	Nominal Value	Mean	COV
b	width	24	0.992*24	0.028
t	thickness	0.5	1.05*0.5	0.044
σ_0	yield stress	34	1.3*34	0.1235
E	Young modulus	29000	0.987*29000	0.076
δ_0	initial deflection	0.35	1.0*0.35	0.05
η	residual stress	5.25	1.0*5.25	0.07

Let us consider a case where we initially have 10 yield stress data. The nominal design value for yield stress of mild steel used here is $\sigma_0 = 34$ ksi. To consider the uncertainty associated with σ_0, we assume a random variable $\hat{\sigma}_0$

as the deviation from the nominal value as $\hat{\sigma}_0 = \sigma_0 - 34$. Hess et al. [1] suggest that the "true" mean yield stress is approximately $\mu_{\sigma_0} \approx 1.3 * 34 = 44.2$ with coefficient of variation 0.1235 and follows a Lognormal distribution. To define this data, we generate 10 random yield stress values from $\sigma_0 = 34 + \hat{\sigma}_0$ with $\hat{\sigma}_0 \sim Lognormal(\mu_{\hat{\sigma}_0} = 1.3 * 34 - 34, \sigma_{\hat{\sigma}_0} = 0.1235 * 1.3 * 34)$, as shown in Fig. 2.2. Obviously, a single probability model form cannot be precisely identified from these data.

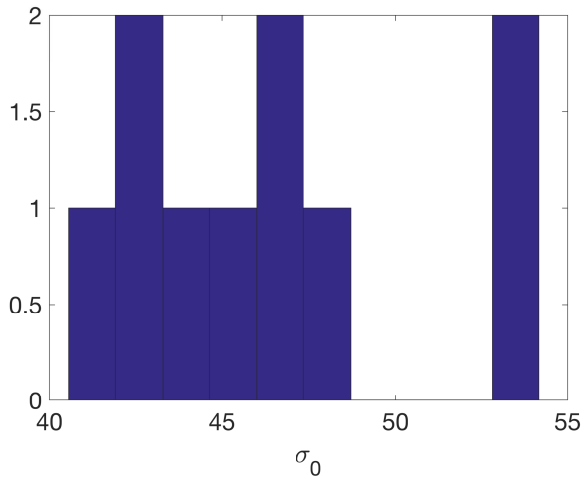


Figure 2.2: Ten randomly generated yield stress values that serve as the initial dataset for uncertainty quantification and propagation in plate buckling strength.

Given the limited data, the next step is to identify the candidate probability models. In this work, 10 candidate models are given, shown in Table 2.2, with the corresponding AIC_c values and model probabilities. Note that the top seven probability models have very small $\Delta_A^{(i)}$ values (similar AIC_c values). In other words, there is almost no difference among the model probabilities for these

seven models. Clearly, it is very difficult to assign a precise "best" model only based on the small data case. Additionally, it should be noted that the last three distribution models have very low probabilities (larger AIC_c values) such that we can remove them from the candidate model set. As a result, the top seven probability models are used to represent this small dataset.

Table 2.2: Ranked candidate probability models based on AIC_c given 10 yield stress values

Rank	Distribution	AIC_c	$\Delta_A^{(i)}$	π_i
1	Inverse Gaussian	61.615	0.000	0.185
2	Lognormal	61.753	0.138	0.173
3	Gamma	61.954	0.338	0.156
4	Log-logistic	62.280	0.664	0.133
5	Rayleigh	62.357	0.742	0.128
6	Nakagmai	62.381	0.765	0.126
7	Weibull	62.956	1.341	0.095
8	Levy	69.952	8.337	0.003
9	Exponential	72.750	11.134	0.001
10	F	98.389	36.774	0.000

Once the probability models are identified, Bayesian inference is applied to estimate the model parameter uncertainty for each specific model form. We assume a noninformative prior that is represented by a uniform distribution with a large range, $U(0, 10^6]$. The joint posterior distributions for six of the models in Table 2.2 are shown in Fig. 2.3. The lower-right and upper-left plots show kernel density estimates of the marginal distributions for each parameter and the lower-left plots presents the contours of the joint posterior density function identified by the MCMC algorithm. We don't show the contour of Rayleigh distribution because it only has one parameter such that the uncertainty is

estimated based on its marginal distribution.

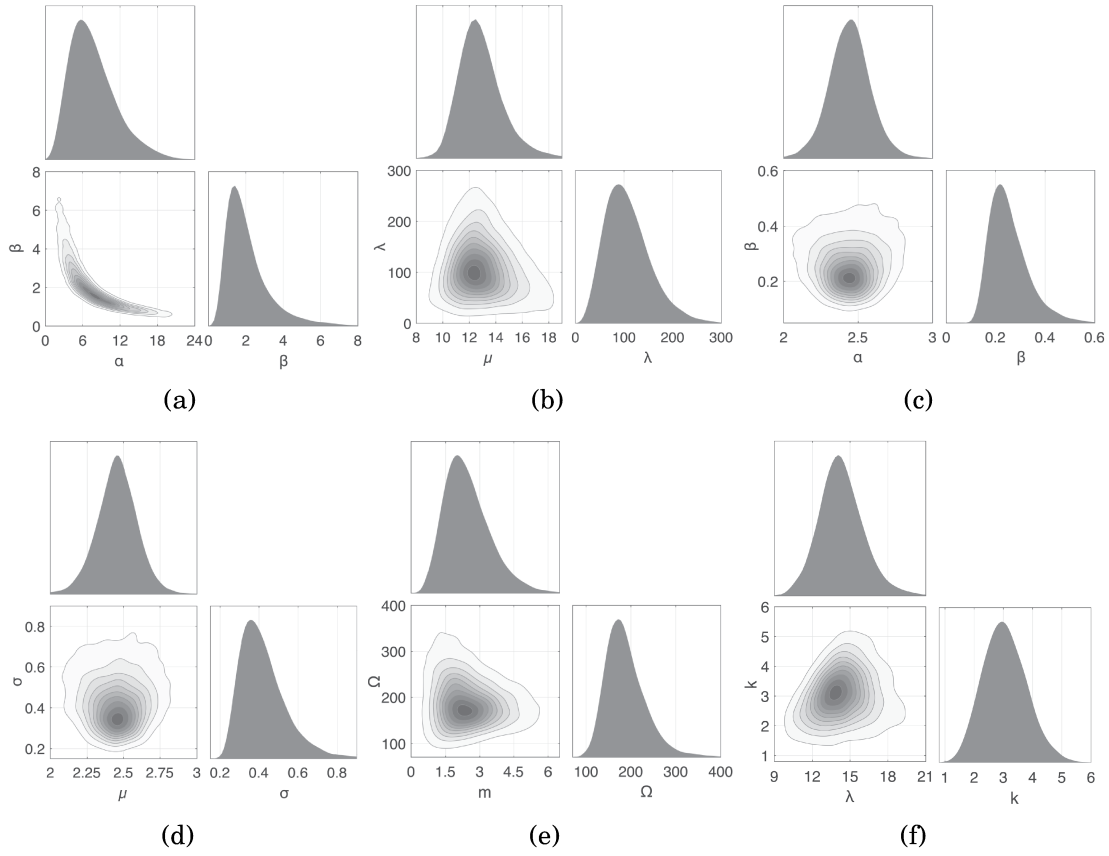


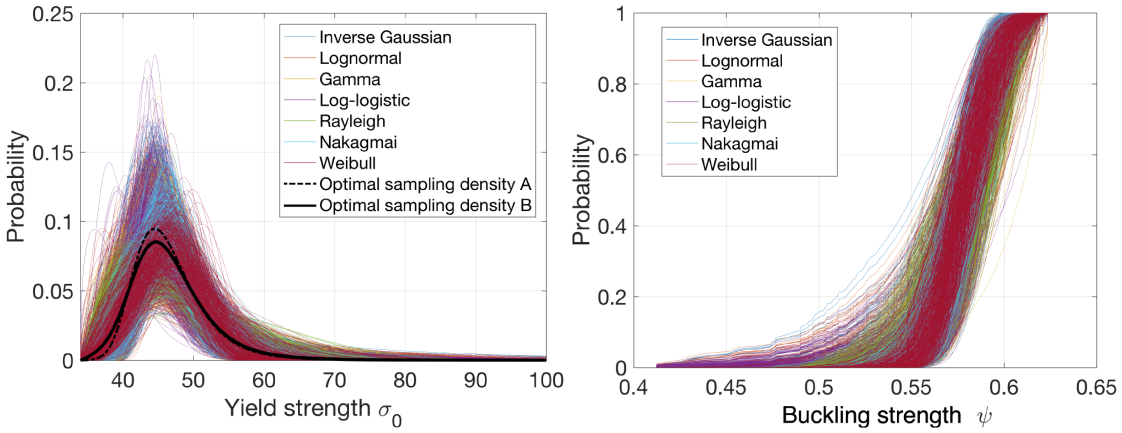
Figure 2.3: MCMC posterior joint parameter densities for the following probability models: (a) Gamma, (b) Inverse Gaussian, (c) Loglogistic, (d) Lognormal, (e) Nakagami, and (f) Weibull.

Using Monte Carlo sampling, the seven models and their joint parameter densities are discretized to obtain a finite set of models. In terms of each sample, the probability model is randomly drawn based on the AIC_c probabilities π_i in Table 2.2 and model parameters are then randomly generated from the joint posterior parameter distribution in Fig. 2.3. Fig. 2.4(a) illustrates the 5000 candidate distributions drawn from the mode set accompanied with two optimal

sampling densities. Optimal sampling density A, $q_A^*(\boldsymbol{x})$ is obtained by minimizing the objective function formulated by the expected Hellinger distance using Eq. (2.34), as shown by the dashed black line in Fig. 2.4(a). Another optimal sampling density B, used in this work, $q_B^*(\boldsymbol{x})$ is identified by minimizing the total expected mean square distance using Eq. (2.42) and is presented as the thick black line in Fig. 2.4(a). Utilizing this optimal sampling density combined with the importance sampling method, the 5000 target densities in Fig. 2.4(a) are efficiently propagated by reweighting 50,000 samples drawn from $q_B^*(\boldsymbol{x})$ according to each of the 5000 target densities. Correspondingly, a cloud of cumulative distribution functions (cdfs) for the buckling strength are shown in Fig. 2.4(b). The response results in Fig. 2.4(b) show the model-form and model parameter uncertainties resulting from the small dataset. Notice that the range of buckling strengths is relatively wide given a fixed probability value.

2.6.1 Separating model-form and parametric uncertainties

This section further explores the effects of model-form and model parameter uncertainties. Fig. 2.4 uses different colors to represent different probability model forms. Thus, the variations within a single color correspond to the parameter uncertainties given a specific probability model form. To further illus-



(a) (b)

Figure 2.4: (a) Candidate pdfs and the optimal sampling density from ten yield stress values, and (b) collection of candidate empirical CDFs for buckling strength ψ.

trate these uncertainties, we may also consider the CDFs of various response statistics based on the different models. The CDFs of the mean and standard deviation of the buckling strength are shown in Fig. 2.5, which presents the conditional CDFs for each probability model and overall CDF considering all probability models. Given a specific probability model form, the conditional CDFs only contain the effects of model parameter uncertainties. The overall CDF is constructed by combining the conditional CDFs with the AIC_c model probabilities in Table 2.2.

Similarly, another investigation aims to study the effects of model parameter and model-form uncertainty on probability of failure. Consider three cases where failure occurs when $\psi_1 < 0.5$, $\psi_2 < 0.55$ and $\psi_3 < 0.6$. Fig. 2.6 shows the empirical CDFs for the probability of failure $P\{\psi_1 < 0.5\}$, $P\{\psi_2 < 0.55\}$ and

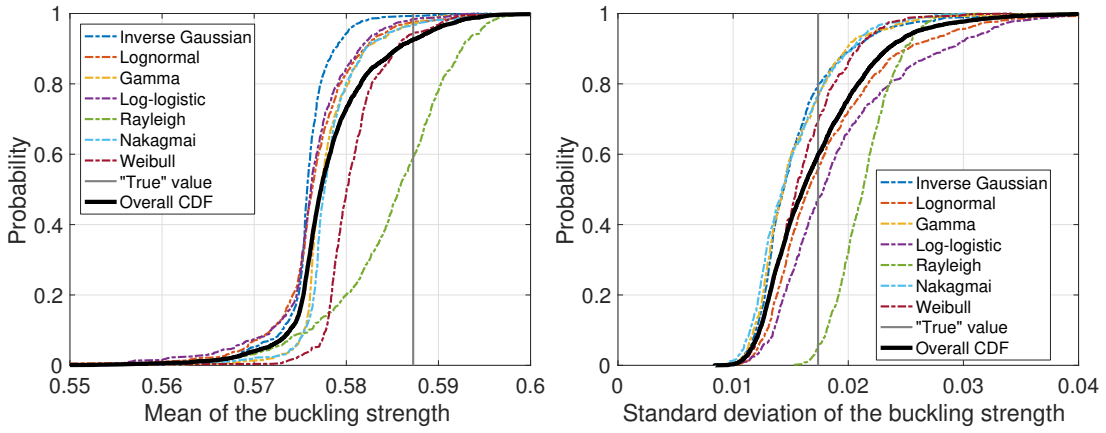
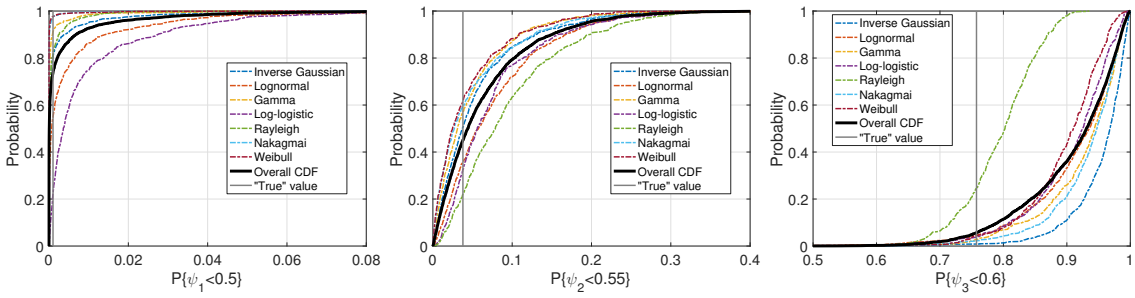


Figure 2.5: Empirical CDFs for (a) mean of buckling strength, and (b) standard deviation of buckling strength ψ.

$P\{\psi_3 < 0.6\}$. Note that these figures also show the conditional CDFs as well as the overall CDFs similar to the case of mean and standard deviation discussed above.

Figs. 2.5 and 2.6 highlight the large uncertainties created by the small dataset. For example, the mean normalized buckling strength ranges from 0.55 to 0.6, while the probability of failure ranges are so large that one cannot place any real confidence in the estimates. Moreover, note that the dominant effects are the model parameter uncertainties, although the model-form uncertainty is important as well as observed in the significant differences between the model specific CDFs in Figs. 2.5 and 2.6.



(a) (b) (c)

Figure 2.6: Empirical CDFs for the probability of failure occurs when (a) $\psi_1 < 0.5$, (b) $\psi_2 < 0.55$, and (c) $\psi_3 < 0.6$.

2.6.2 Effect of dataset size

In this section, we investigate the convergence of the buckling strength as a function of dataset size. As discussed in the previous section, small datasets led to very large uncertainties including model-form and model parameter in the buckling strength. This raises an important question: "How much data is necessary to gain adequate confidence in the buckling strength probabilities?"

Fig. 2.7 shows the AIC_c probabilities for each candidate probability model as a function of dataset size. Until 1000 measured yield stress data, the multi-model inference does not select the single "correct" Lognormal model. In other words, model-form uncertainty plays an important role prior to the collection of a very large dataset. This also begs another question: "How does this influence the uncertainty in buckling strength?".

The evolution of different quantities for various dataset sizes are systematically shown in Figs 2.8-2.12. Fig. 2.8 illustrates the sets of candidate prob-

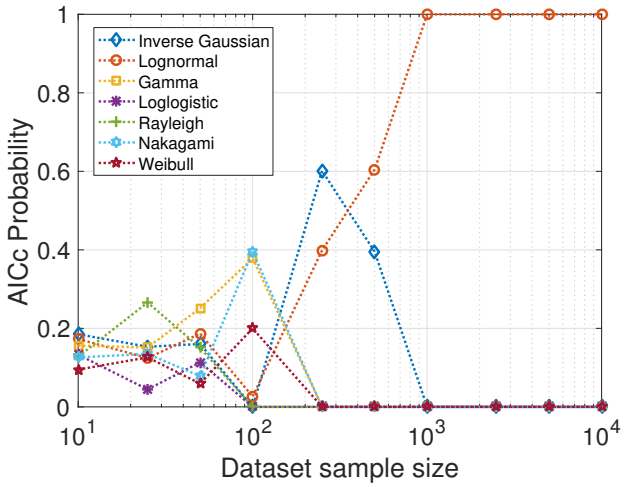


Figure 2.7: AIC_c probability as a function of dataset size.

ability densities along with the initial optimal sampling density derived from only 10 data for giving datasets. Even though the optimal sampling density loses optimality as data are collected, the optimal sampling density can be still used for propagation of uncertainties at the potential expense of increased variance in statistical estimates. We see from Fig. 2.8 that the cloud of candidate densities gradually narrows as data are further added as expected. The corresponding set of CDFs for the buckling strength are shown in Fig. 2.9. One important feature of the methodology is that no additional model evaluations were required to estimate these CDFs. We only need to reweight the 50,000 samples drawn from the original optimal sampling density to reflect the updated candidate densities. As expected, the cloud of response CDFs gradually narrows toward the "true" estimate as more data are collected.

Fig. 2.10-2.12 show statistical estimates of the mean, standard deviation

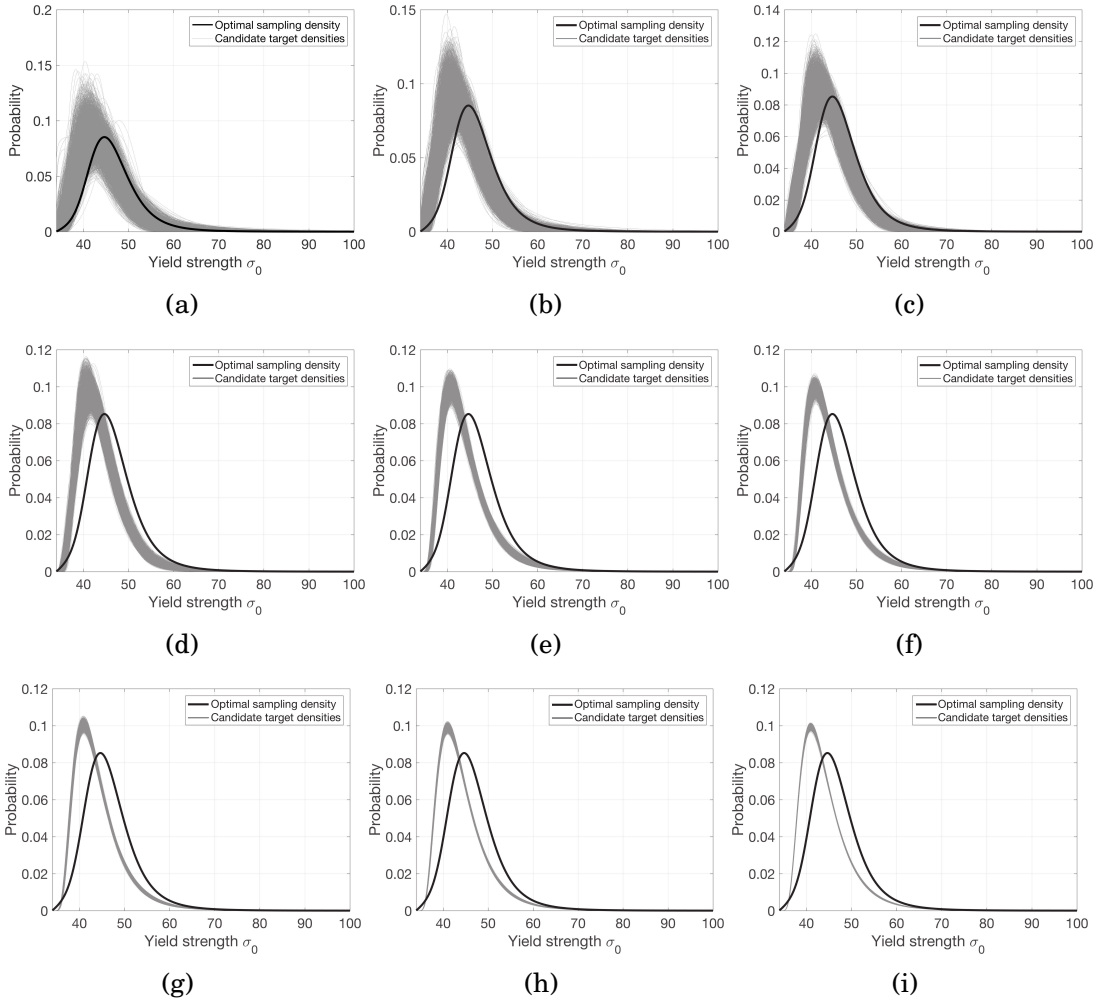


Figure 2.8: Optimal sampling density with candidate target densities based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 250 data, (e) 500 data, (f) 1000 data, (g) 2500 data, (h) 5000 data and (j) 10000 data.

and probability of failure $P\{\psi < 0.5\}$ CDFs for increasing dataset sizes. For each candidate model, the conditional CDFs and overall CDFs are both shown along with the "true" values estimated by Monte Carlo method with large samples (10^6 used here) drawn from the "true" Lognormal distribution. Notice that the model-form uncertainty for the mean and standard deviation becomes in-

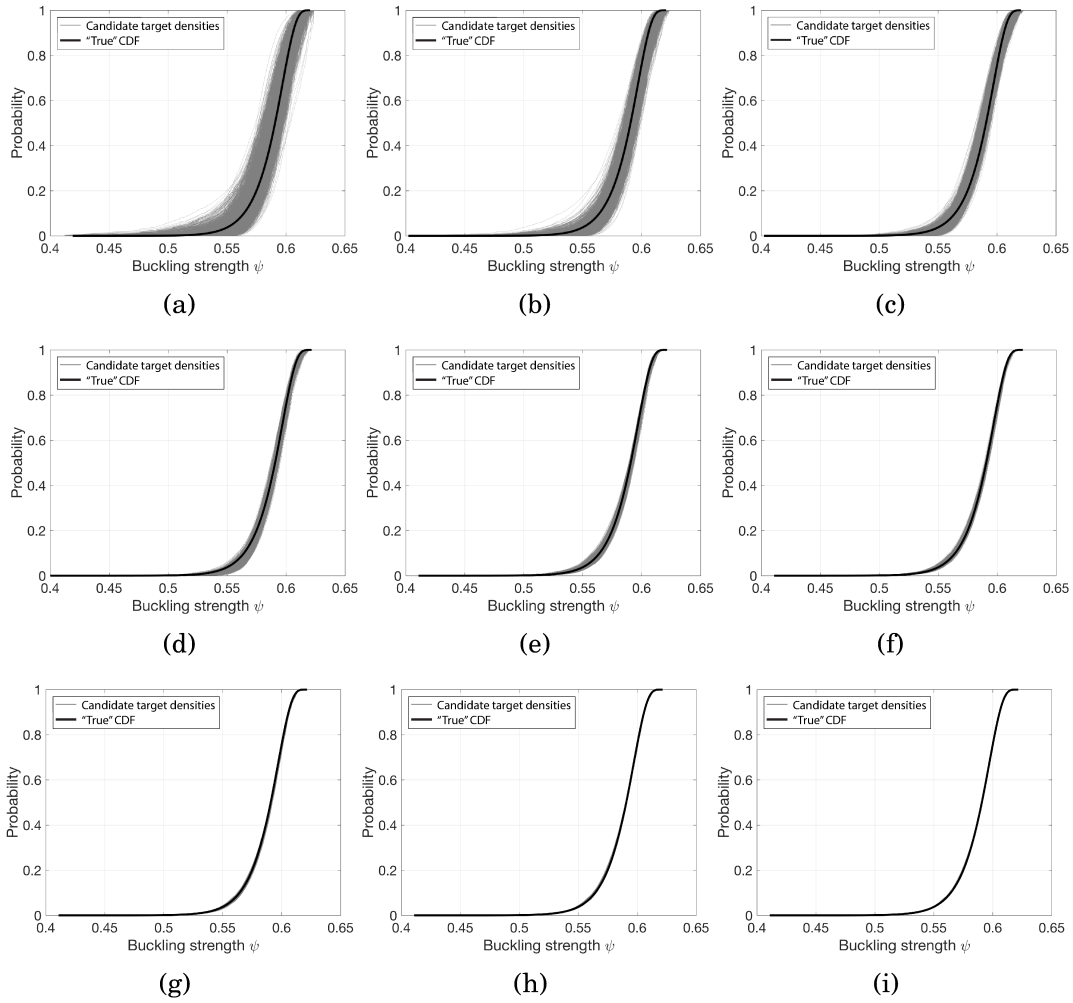


Figure 2.9: CDFs for the buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 250 data, (e) 500 data, (f) 1000 data, (g) 2500 data, (h) 5000 data and (j) 10000 data.

significant after approximate 100 data, while model-form uncertainty remains significant for the probability of failure until approximate 1000 data are collected, at which point the single correct Lognormal distribution is identified (Inverse Gaussian model can be eliminated).

Model parameter uncertainty constitutes the greater proportion of the total

uncertainty and is basically reflected by the breadth of the CDFs of the statistical quantities as shown in Fig. 2.10-2.12. A lack of model parameter uncertainty corresponds to a CDF which is a simple step function. We also note that the model parameter uncertainty diminishes a little more slowly than the model-form uncertainty and still remains significant even for very large dataset size. But as expected, the CDFs still gradually converge toward the true estimate values as the dataset size increases to be very large.

2.6.3 Convergence analysis

For decision makers, it is often more important to identify a confidence threshold and thus determine how much data must be collected to meet this threshold. Alternatively, one can also determine the amount of data that can be feasible collected and estimate the confidence level that can be achieved given this dataset size. In this work, we define a simple confidence metric as the range of the upper and lower quantiles of width 0.025 for the statistical quantity Y considering model-form and model parameter uncertainties given n data as follows:

$$\delta_Y^{(n)} = Q_{0.975}(Y^{(n)}) - Q_{0.025}(Y^{(n)}) \tag{2.47}$$

We denote $\delta_\mu^{(n)}$, $\delta_\sigma^{(n)}$, and $\delta_{(\psi<\psi^*)}^{(n)}$ as the ranges for the mean, standard deviation and probability of failure. The relationship between these bounds and dataset

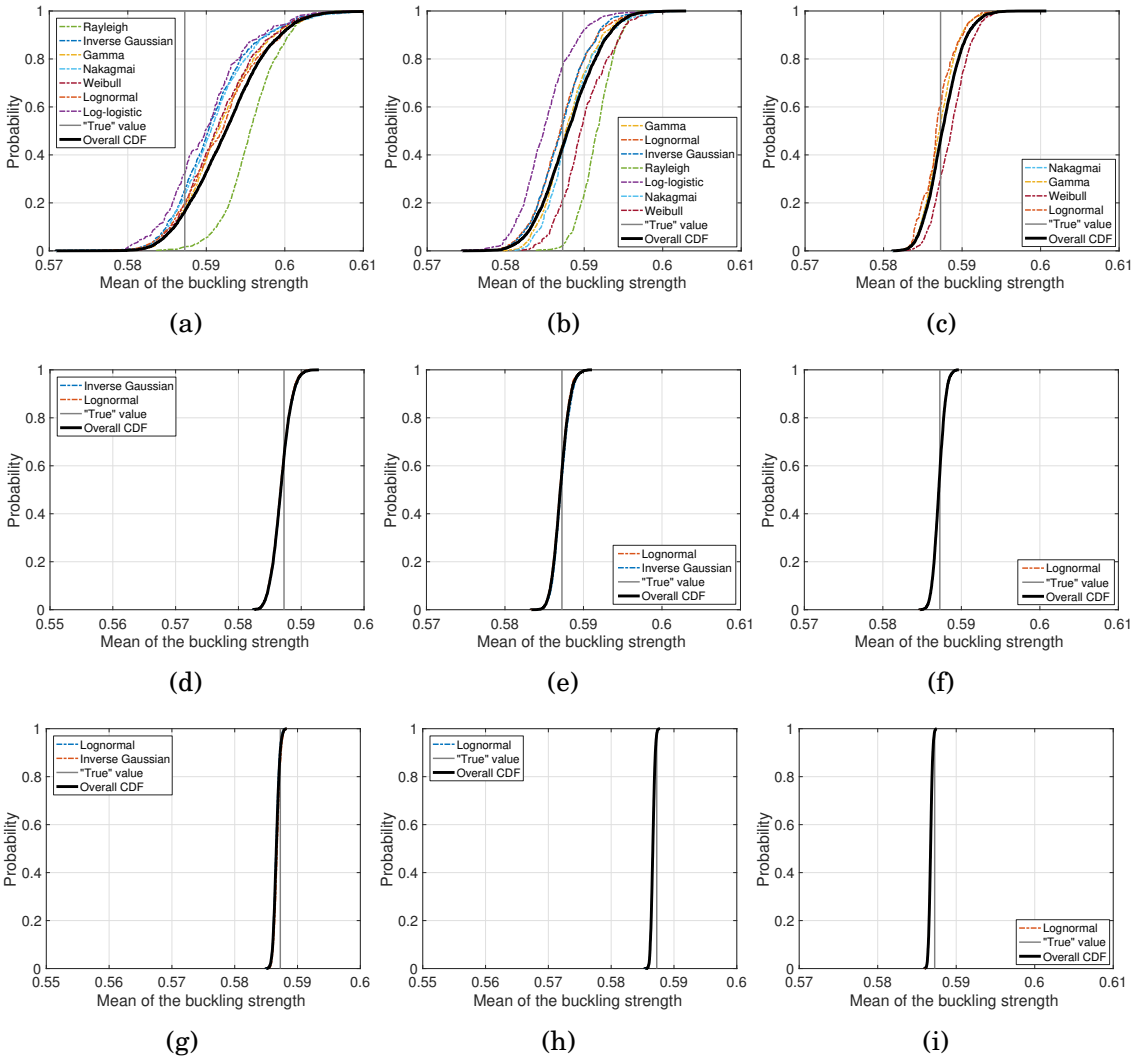


Figure 2.10: CDFs for the mean buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 250 data, (e) 500 data, (f) 1000 data, (g) 2500 data, (h) 5000 data and (j) 10000 data.

size is shown in Fig. 2.13. For decision makers, these plots may be very useful as they provide easy-to-interpret information in determining the appropriate amount of data to collect and the confidence that can be achieved. Notice that the probability ranges in these plots for the various statistics converge at a rage of $\sim \frac{1}{n^2}$. It will be helpful to make prediction and decisions if this is found

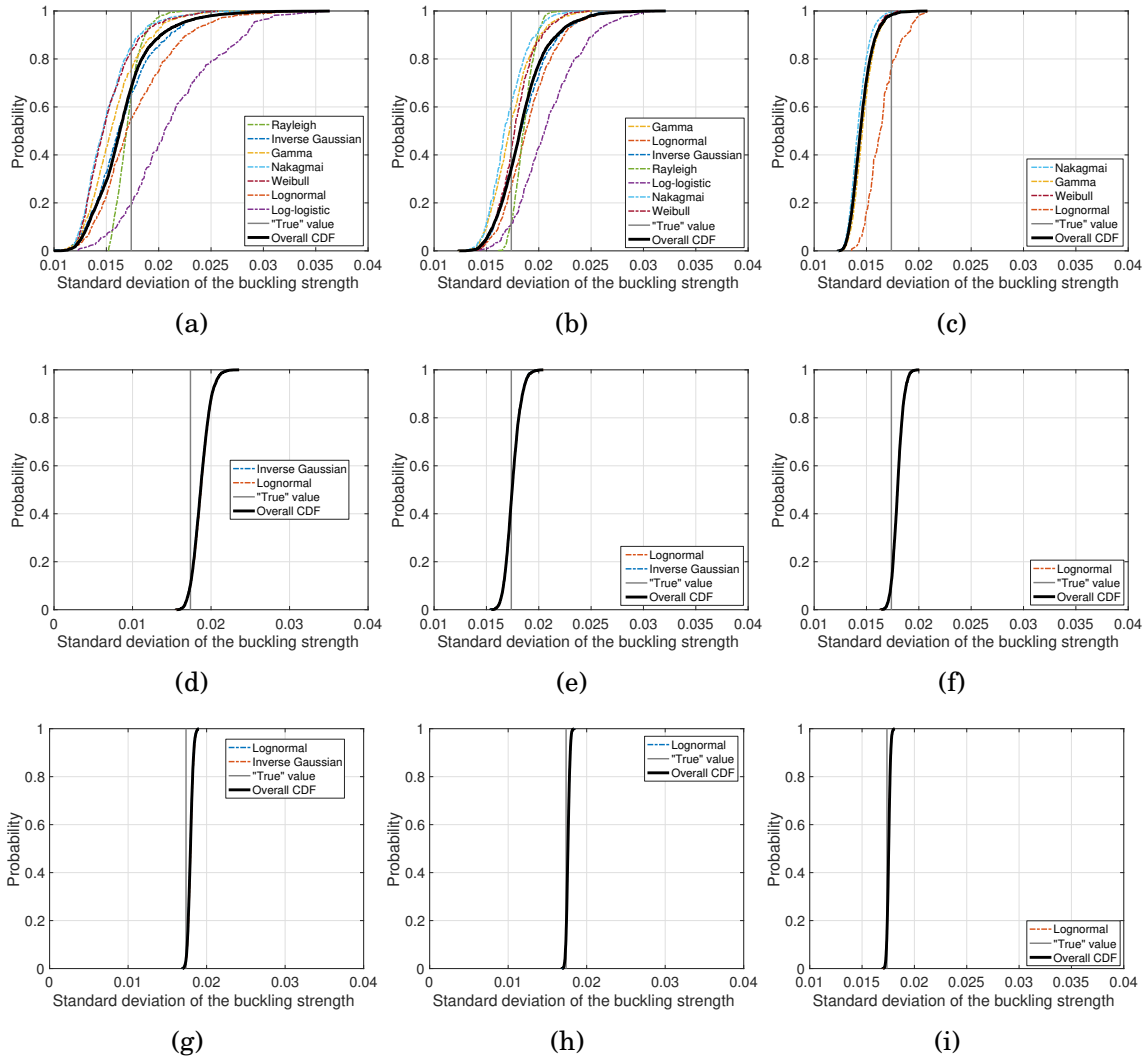


Figure 2.11: CDFs for the standard deviation of the buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 250 data, (e) 500 data, (f) 1000 data, (g) 2500 data, (h) 5000 data and (j) 10000 data.

to extend more generally. We may therefore collect small data initially and then project the confidence level, which will be achievable to guide how much additional data to collect in order to meet a specified uncertainty tolerance. The quantile range used here is simply for illustration and the plots of this nature can be straightforward to generalize for other measures.

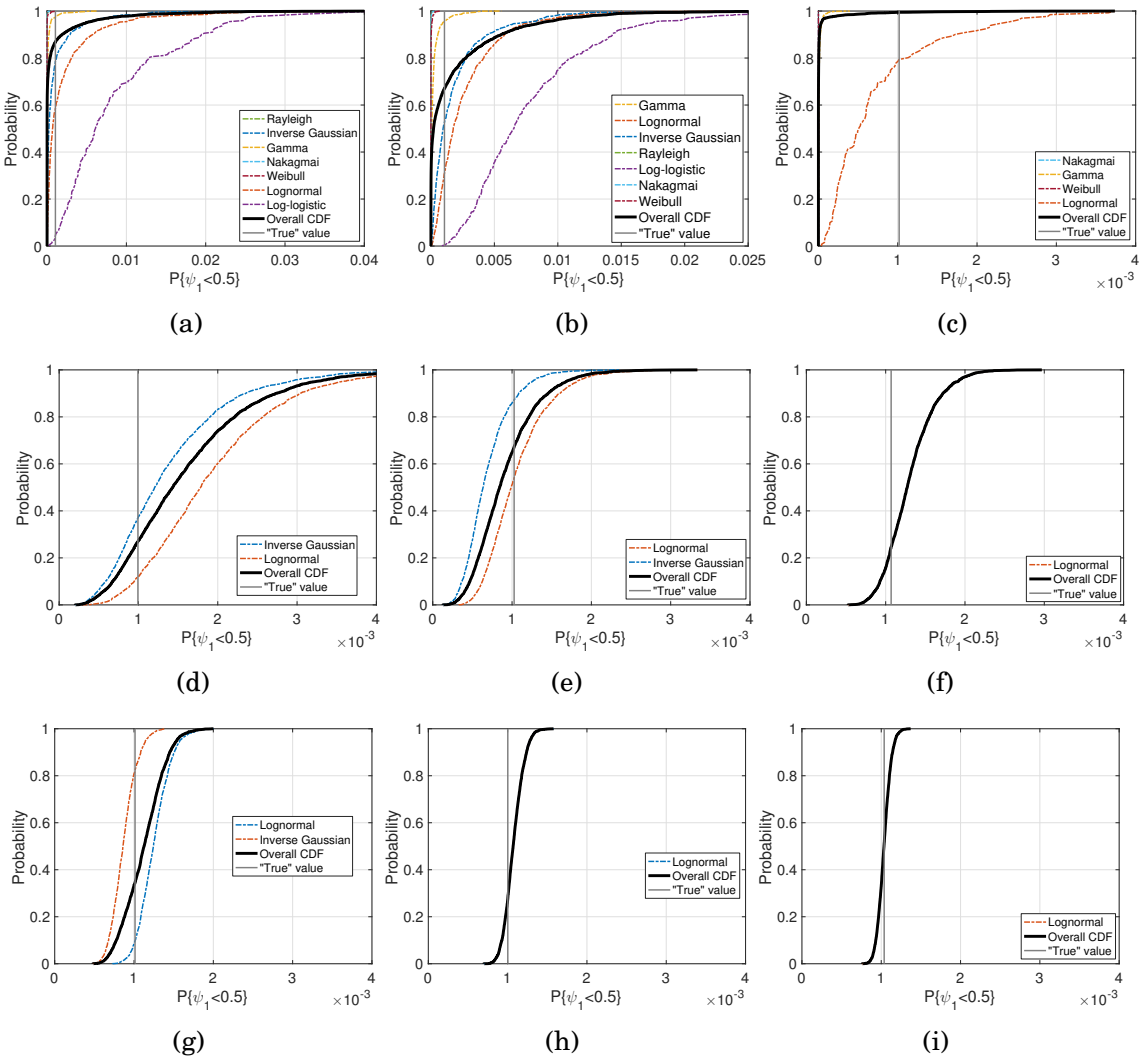
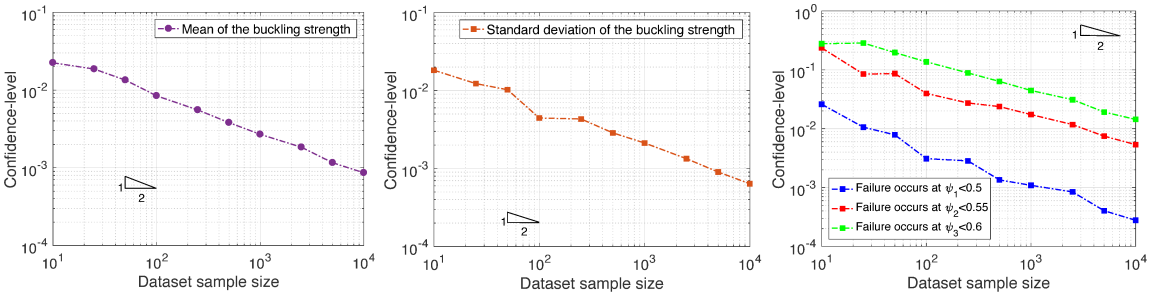


Figure 2.12: CDFs for the probability of failure $P\{\psi < 0.5\}$ based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 250 data, (e) 500 data, (f) 1000 data, (g) 2500 data, (h) 5000 data and (j) 10000 data.

2.7 Conclusion

A novel methodology has been proposed to quantify and propagate uncertainty resulting from small datasets. The methodology adopts the concepts of multimodel inference to identify a set of candidate probability models which



(a) (b) (c)

Figure 2.13: Convergence of the probability range for (a) mean, (b) standard deviation, and (c) probability of failure

are representative of the given small dataset. A model probability is assigned for each of these candidate probability models using either information-theoretic or Bayesian multimodel inference. Given a specified plausible model, the model parameter uncertainty, quantified by the joint posterior probability density of parameters is estimated by Bayesian inference. Therefore, the candidate probability models along with the joint parameter densities quantify the total epistemic uncertainty created by the small dataset.

The proposed methodology differs from the prior methods that reduce the full probabilistic description to a single probability model using averaging or other approaches. Instead, we herein retain and propagate all candidate models with their joint parameter densities. This is achieved by utilizing the concept of importance sampling. An optimal sampling density identified through an analytical optimization method best fits the full suite of probability models and is propagated using Monte Carlo simulation and the samples are reweighted

based on each of the candidate probability models. A significant advantage of the proposed methodology is to reduce a multi-loop Monte Carlo with n^3 samples to a single-loop Monte Carlo with n samples.

The proposed methodology effectively treats uncertainties in the context of small datasets. The effect of small dataset size is also fully investigated. When datasets are very small, i.e. $N < 100$, both model-form and model parameter uncertainties are very large and the statistical quantities also show large variabilities. In other words, there is a lack of confidence in the response. With the increasing of dataset size N, we show that the uncertainties narrow and response converge toward their true probabilities.

Chapter 3

The effect of prior probabilities on uncertainty quantification and propagation

This chapter presents an investigation into the effect of prior probabilities on the uncertainties that results from the multimodel uncertainty quantification and propagation method presented in Chapter 2. The UQ methodology employed in this chapter follows the fully Bayesian framework presented in Chapter 2 that enables larger flexibility in studying the effect of prior probabilities.

When dealing with small datasets, prior probabilities, particularly informative priors, in both model-form and model parameters have a significant

influence on uncertainty quantification and propagation through a computational model. These effects are systematically studied for the same plate buckling strength problem described in Chapter 2, with uncertainties in material properties. It is shown that prior probabilities can play a very important role in multimodel UQ for small datasets and inappropriate priors may even have lingering effects, which yiels biased estimates even for large dataset size. In terms of uncertainty propagation, this effect may lead to inaccurate or wrong probability bounds on response outputs.

3.1 Formulating model and parameter priors

Given a set of probability models $\mathcal{M}$, the implementation of Bayesian multimodel inference strongly depends on the specification of the prior model probabilities $p(M_j)$ and prior parameter probabilities $p(\boldsymbol{\theta}_j|M_j)$ for each model M_j. Generally speaking, a reasonable choice of the prior may have only a minor impact on the posterior estimation given large datasets. But if only limited data is available or prior information is not entirely appropriate, the choice of prior will play a very important role. This chapter first briefly reviews some critical concepts including noninformative and informative priors, and then discusses how to formulate data-driven, informative priors for Bayesian multimodel in-

ference.

3.1.1 Parameter prior probabilities

Prior probabilities for model parameters have a substantial influence in Bayesian multimodel uncertainty quantification and propagation. Noninformative and informative priors are broadly distinguished here and we also investigate how these different priors can be built under various conditions, i.e. no prior information is available, existing historical data and subjective assumptions.

3.1.1.1 Noninformative priors

Generally speaking, a uniform prior, that is diffuse, flat and often referred to as a vague prior, is one of the most common noninformative priors. But a vague or diffuse prior is not necessarily uniform and sometimes a diffuse prior can be more informative than the uniform prior [121–123]. Usually, the uniform prior can be given as

$$p(\boldsymbol{\theta}_j|M_j) = \text{Constant}, \quad \boldsymbol{\theta}_j \in \Omega_{\theta_j} \quad 1 \leq j \leq m \tag{3.1}$$

where the range of $\boldsymbol{\theta}_j$, Ω_{θ_j} is a subset of the parameter space Θ_j ($\Omega_{\theta_j} \subset \Theta_j$). This indicates that there is no *a priori* reason to favor any particular param-

eter value. Instead, we only know its range $\boldsymbol{\theta}_j \in \Omega_\theta$. Therefore, the posterior distribution is proportional to the likelihood [124]

$$p(\boldsymbol{\theta}_j|\boldsymbol{d}, M_j) \propto p(\boldsymbol{d}|\boldsymbol{\theta}_j, M_j), \quad \boldsymbol{\theta}_j \in \Omega_{\theta_j} \quad 1 \leq j \leq m \tag{3.2}$$

If the range Ω_{θ_j} is specified as the parameter space $\Omega_{\theta_j} = \Theta_j$, Bayesian inference under a flat prior may lead to an improper prior if

$$\int_{\Theta_j} p(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j = \infty, \quad 1 \leq j \leq m \tag{3.3}$$

The normalizing constant in this case may not exist. One needs to ensure that the posterior is proper when an improper prior is used.

Another commonly used noninformative prior, the Jeffreys prior [125] is defined to be proportional to the square root of the determinant of the Fisher information matrix

$$p(\boldsymbol{\theta}_j|\boldsymbol{d}, M_j) \propto |J(\boldsymbol{\theta}_j)|^{1/2}, \quad 1 \leq j \leq m \tag{3.4}$$

where $J(\cdot)$ is referred to as the Fisher information which is given as

$$J(\boldsymbol{\theta}_j) = -\int E\left[\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\theta}_j, M_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j^T}\right] p(\boldsymbol{x}|\boldsymbol{\theta}_j, M_j) d\boldsymbol{x}, \quad 1 \leq j \leq m \tag{3.5}$$

For certain models, the Jeffreys prior is often an improper prior as it cannot be normalized.

In this study, we apply proper uniform priors as representative noninformative priors. The ultimate intention here is to investigate the effect of a suitably representative noninformative prior on multimodel uncertainty quantification and propagation against the effects of different informative priors.

3.1.1.2 Informative priors

Differing from the noninformative priors discussed previously, informative priors yield a posterior that is not dominated by the likelihood function; on the contrary, informative priors play a fundamental role on the posterior distribution. This is particularly true for inference in the case of small datasets. The appropriate use of informative priors shows the power of the Bayesian methodology: information or knowledge gathered from past experience, previous studies or expert opinions can be combined with additional data in a natural way. As a result, the informative prior can be interpreted as the state of subjective prior knowledge. Nonetheless, in practice, it is often intractable to assign an informative prior precisely and historical experiments, experience and data may not be completely appropriate for the current situation. Consequently, the prior specification based on subjective knowledge may not be unbiased. One principal objective of this work is to fully understand the effect of such imprecise

and/or incorrect informative priors on multimodel uncertainty quantification and propagation with small datasets.

To avoid formulating priors according to assumptions or intuition, this study aims to formulate data-driven informative priors by exploiting historical data, denoted $\hat{\boldsymbol{d}}$, as may be available in the literature. For such cases, the historical data $\hat{\boldsymbol{d}}$ represents the existing state of knowledge as objectively as possible. However, these data may not be entirely appropriate for the current problem and hence may or may not provide "good" priors.

The data-driven prior is quantified by applying Bayes' rule to the historical data, $\hat{\boldsymbol{d}}$. Using the currently observed data, $\boldsymbol{d}$, with a non-informative pre-prior, the posterior then becomes the prior for Bayesian inference on the currently observed data, $\boldsymbol{d}$. A suitable noninformative prior, named the "pre-prior", is employed in the initial Bayesian inference. Under this framework, the currently observed data $\boldsymbol{d}$ is effectively treated as an extension of the historical data $\hat{\boldsymbol{d}}$. If the historical dataset is small, the resulting prior is referred to as weakly informative and retains some influence of the noninformative pre-prior. If the historical dataset is relatively large, the resulting prior is referred to as strongly informative and dominates the pre-prior.

The algorithm used in this work for the defined informative priors can be summarized in the following three stages:

- *Stage 1: Noninformative pre-prior* - Noninformative pre-priors $\hat{p}(\boldsymbol{\theta}_j|M_j)$

can be constructed in many different ways. Given the likelihood function $\hat{p}(\hat{\boldsymbol{d}}|\boldsymbol{\theta}_j, M_j)$, one may derive the noninformative prior according to Jeffrey's rule, or use a flat prior (uniform prior) instead. Here, we use a uniform pre-prior.

- *Stage 2: Pre-Bayesian inference* - A pre-Bayesian inference is adopted to estimate the the posterior distribution based on the historical data $\hat{\boldsymbol{d}}$ combined with a given noninformative prior $\hat{p}(\boldsymbol{\theta}_j|M_j)$ and the specified model M_j

$$p^*(\boldsymbol{\theta}_j|M_j) = \hat{p}(\boldsymbol{\theta}_j|\hat{\boldsymbol{d}}, M_j) = \frac{\hat{p}(\hat{\boldsymbol{d}}|\boldsymbol{\theta}_j, M_j)\hat{p}(\boldsymbol{\theta}_j|M_j)}{\int_{\Theta_j} \hat{p}(\hat{\boldsymbol{d}}|\boldsymbol{\theta}_j, M_j)\hat{p}(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j}, \quad 1 \leq j \leq m \quad (3.6)$$

The posterior distribution $\hat{p}(\boldsymbol{\theta}_j|\hat{\boldsymbol{d}}, M_j)$ is selected as the prior probability $p^*(\boldsymbol{\theta}_j|M_j)$ for the currently observed data $\boldsymbol{d}$.

- *Stage 3: Nonparametric estimate from posterior samples* - Eq. (3.6) is typically solved implicitly using an MCMC algorithm. Thus, the data-driven prior is not available in closed-form for Bayesian updating using the new additional data, $\boldsymbol{d}$. Hence, a nonparametric kernel density estimate is employed to approximate the unknown prior probability density distribution from the posterior samples using MCMC algorithm.

  Regarding the multivariate density functions involving parameter vector

$\boldsymbol{\theta}_j$ with dimension K_j, the kernel density estimate $\tilde{f}(\boldsymbol{\theta}_j|M_j)$ has the form

$$\tilde{f}(\boldsymbol{\theta}_j|M_j) = \frac{1}{n}\sum_{k=1}^{n}\prod_{i=1}^{K_j}\left\{\frac{1}{w_i}\phi\left(\frac{\boldsymbol{\theta}_{j,i}-\boldsymbol{\theta}_{j,i}^k}{w_i}\right)\right\}, \quad 1 \leq j \leq m \tag{3.7}$$

given a sample set $\boldsymbol{\theta}_j = \left\{\boldsymbol{\theta}_j^1, \boldsymbol{\theta}_j^2, ..., \boldsymbol{\theta}_j^n\right\}$ of size n, a model M_j [126] and $\boldsymbol{\theta}_{j,i}^k$ is the k^{th} sample in the i^{th} dimension of $\boldsymbol{\theta}_j$, and w_i is the corresponding bandwidth. $\phi(\cdot)$ is a chosen as a Gaussian kernel given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \tag{3.8}$$

The kernel bandwidth is then determined by minimizing the asymptotic mean integrated square error (AMISE) [127] such that, for the Gaussian kernel, the optimal bandwidth is

$$w_i^{opt} = \left[\frac{4}{K_j+2}\right]^{1/(K_j+4)} n^{-1/(K_j+4)}\sigma_i \tag{3.9}$$

where σ_i is the standard deviation of the samples $\left\{\boldsymbol{\theta}_{j,i}^1, \boldsymbol{\theta}_{j,i}^2, ..., \boldsymbol{\theta}_{j,i}^n\right\}$. The kernel density estimate $\tilde{f}(\boldsymbol{\theta}_j|M_j)$ is then used as the informative prior for Bayesian inference on the observed data $\boldsymbol{d}$.

3.1.2 Prior model probabilities

Prior model probabilities also have an essential impact on the multimodel uncertainty quantification and propagation. The concept of noninformative and informative parameter prior can be also extended to the prior model probabilities.

A simple and widely used choice for the prior model probability $p(M_j), j = 1, ..., m$, is the uniform prior

$$\pi_j = p(M_j) = \frac{1}{m} \tag{3.10}$$

This prior can be regarded as noninformative in the sense of favoring all models equally. Using this prior, the posterior model probability is equal to the ratio of the model evidence to the cumulative evidence,

$$\hat{\pi}_j = p(M_j|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|M_j)}{\sum_{k=1}^{m} p(\boldsymbol{d}|M_k)} \tag{3.11}$$

These asymptotically correspond to the BIC model probabilities as discussed previously, but the apparent noninformativeness of Eq. (3.11) can be deceptive. This is because the model prior is only uniform in probability and will typically not be uniform on the model characteristics. As a result, given that several models are very similar and only a few are different, the posterior model prob-

abilities in Eq. (3.11) may be biased in such a way that they do not favor the accurate models [128].

In practical applications, model prior probabilities are usually chosen based on subjective assumptions or preferences that may be obtained from expert opinion, previous experience or historical data. It is extremely important because strong prior beliefs can considerably influence posterior model probabilities that can cause very inaccurate (if the priors are incorrect) or very accurate (if the priors are correct) assessments of uncertainty. In this work, we take account of these respective prior model probabilities and aim at understanding their influence on multimodel uncertainty quantification and propagation.

3.2 Application to plate buckling strength problem

In this work, we investigate the effect of model and parameter priors on buckling strength problem described in Section 2.6. The design buckling strength is based on nominal values for the six variables in Eq. (2.46) provided in Table 2.1. Similarly, we place the emphasis on assessing the influence of uncertainty in the yield strength σ_0 and investigate the effect of prior on multimodel uncertainty quantification and propagation.

3.2.1 Description of historical data

Hess et al. [1] provided a review of uncertainties in geometric and material properties of structural steel for ship building applications. They collected a series of test/measurement data from reports by the Ship Structure Committee (SSC) [129, 130] and conducted a systematical statistical analysis based on these data. As part of an effort, they also established a database of marine steel properties and tests/measurements performed by the Naval Surface Warfare Center, Carderock Division (NSWCCD). Although these past sources of yield strength data are sparse, they may play an important role since they provide a valuable source of prior information for design and manufacturing of ship components. As the data is limited, it is difficult to represent the uncertainties associated with the design variable. Therefore, it is highly necessary to quantify the uncertainties and variations in these material properties. This work will utilize the historical experimental data to estimate uncertainty in the yield strength of mild steel. These material property data are collected from a number of historical reports including SSC-142 [131], SSC-145 [132] and SSC-352 [133]. There are four classes of structural steel included in these datasets:

- *ABS-A* - plates with thickness not exceeding 1/2 inch and all shapes
- *ABS-B* - plates with thickness over 1/2 inch but not exceeding 1 inch

- *ABS-C* - plates with thickness over 1 inch
- *ASTM-A7* - Historical conventional structural steel alloy replaced by ASTM-A36

Typically, these three ABS steels are used for ship-building and marine steels, and possess nominally the same design properties (most notably $\sigma_0 = 34$ ksi) while they vary somewhat in chemical composition. However, the ASTM-A7 is a historical carbon steel having design yield strength in the range $\sigma_0 = [30, 33]$ ksi. The work of Hess et al. [1] presented a statistical analysis of these yield strength data, summarized in Table 3.1. These data are representative of the type of historical data that may be available (these tests data back to 1948) and from existing literature very useful for our investigation because they can be used for assigning prior distributions in Bayesian inference but are not necessarily representative of what may be expected from modern materials. As a result, the statistical analysis of the four materials in Hess et al. [1] provides various priors from which to initiate our investigation.

Table 3.1: Statistical information and comments of informative knowledge from historical data, summarized from [1].

Steel type	Min	Max	Mean	COV	Distribution	# of tests	Comments
ABS-A	31.9	39.6	36.091	0.059	Lognormal	33	Weakly informative but incorrect
ABS-B	27.6	46.8	34.782	0.116	Lognormal	79	Informative and correct
ABS-C	30.9	41.5	33.831	0.081	Lognormal	13	Weakly informative but incorrect
ASTM-A7	28.6	49.4	38.197	0.108	Normal	58	Informative but incorrect

Since the ship structural plate with thickness $t = 0.75$ inch is our interest

here, we assume that the "true" model is ABS-B material class given by Table 3.1. In fact, the model given in Table 3.1 is not likely the true model for ABS-B material but for our purposes it is useful to consider it as true to provide a baseline from which we have an *informative* and *correct* prior. The ABS-A and ABS-C materials are similar to the "true" ABS-B material and their datasets are smaller. Consequently, they are taken as weakly informative but technically incorrect priors. The ASTM-A7 significantly differs from the other three materials. We therefore consider it as an informative but incorrect prior due to its comparatively large dataset. It is worth noting that an analyst may consider any one of these data sets to be "close enough" in order to define a prior for UQ (justifiably or not) under practical conditions of small data. The main objective of this work is to investigate the impact of using these different priors in the context of multimodel Bayesian uncertainty quantification and propagation.

Fig. 3.1 presents histograms of the material data for ABS-A, ABS-B, ABS-C and ASTM-A7. The ABS-B material data collected from the technical report SSC-142 [131] is assumed to follow a Lognormal distribution with mean $\mu = 34.782$ and with coefficient of variation 0.116. Again, we assume this to be the "true" model and all "data" are synthetically generated from $\sigma_0 \sim \text{Lognormal}(\mu_{\sigma_0} = 34.782, \sigma_{\sigma_0} = 0.116 * 34.782)$. Fig. 3.2 shows a histogram of 10 simulated yield strength data. A unique probability model cannot be precisely

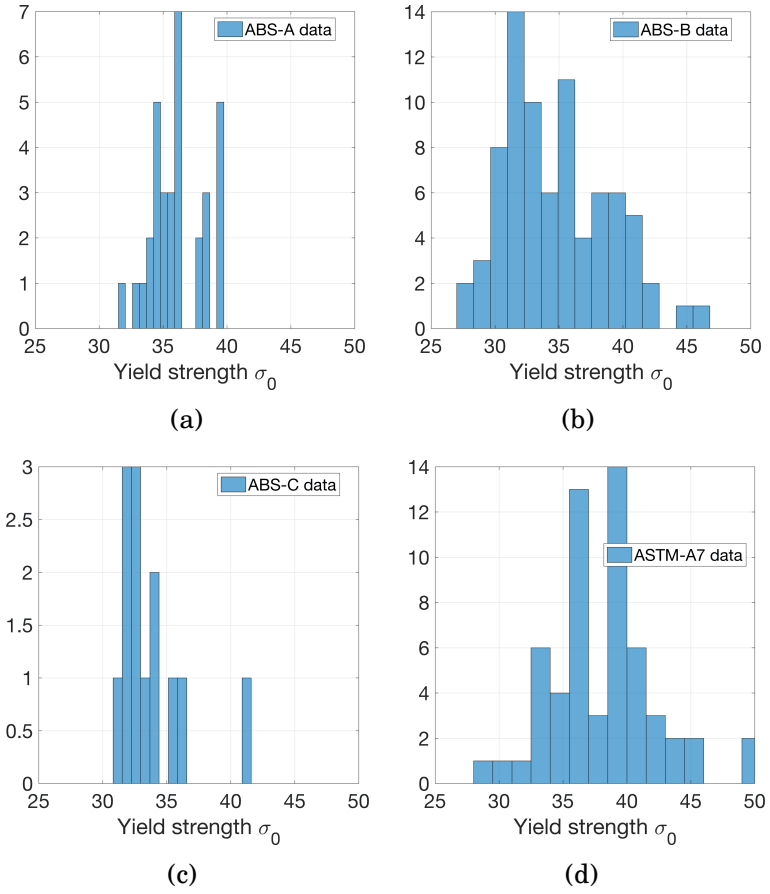


Figure 3.1: Histograms of material data for (a) ABS-A, (b) ABS-B, (c) ABS-C and (d) ASTM-A7

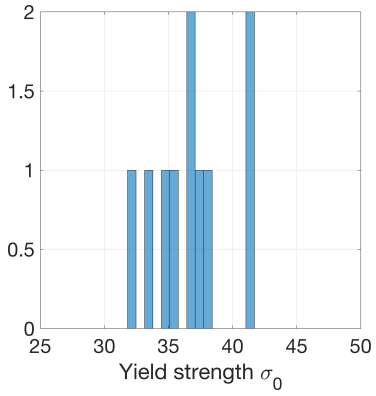


Figure 3.2: Ten randomly sampled yield strength data that serve as the initial dataset

assigned if only 10 data and same prior data are given in this case. Instead, the following seven candidate probability models are considered herein: Gamma, Inverse Gaussian, Logistic, Loglogistic, Lognormal, Normal and Weibull. Regarding each of these models, we derive the prior parameter densities using the dataset in Fig. 3.1 and the method in Section 3.1.1.2.

3.2.2 Influence of data-driven priors on uncertainty quantification

Within the Bayesian multimodel framework, there are two stages of inference related to model-form uncertainty and model parameter uncertainty. The flowchart in Fig. 3.3 shows an interesting interplay between these two stages of inference as follows:

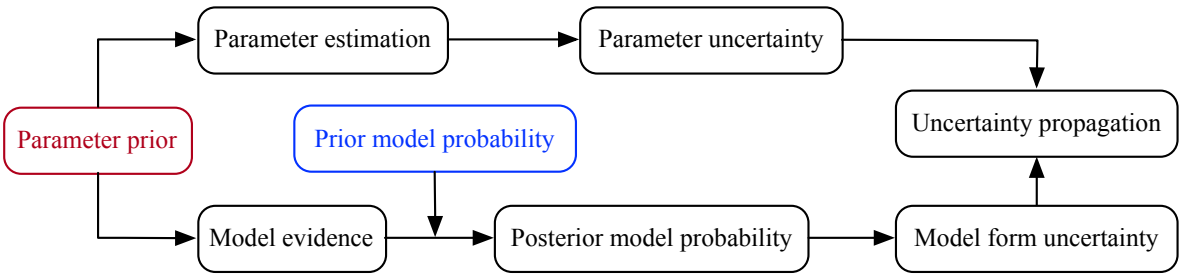


Figure 3.3: Influence of parameter prior and model prior probability on uncertainty quantification and propagation

Multiple candidate models, such as the seven model listed above, are considered in the first stage. Informed by expert opinion, some assumptions are

made with respect to their prior model probabilities. According to Bayes' rule (Eqs. (2.11)-(2.12)), the model probabilities are updated as more data are collected. Also, the selection of the parameter prior has an impact on the these updated probabilities. In the second stage, for each model form, the model parameter distribution is estimated by Bayesian inference based on the data. They obviously rely on the prior parameter probabilities. These processes are employed to provide the posterior estimation used to quantify uncertainty in the parameter of interest (here σ_0).

3.2.2.1 Effect of priors on model-form uncertainty

General speaking, it is common to assign equal prior probability (i.e. $\pi_j = P(M_j) = 1/m = 1/7$ [128] in Bayesian model selection. For some cases, subjective non-equal probabilities may be assumed. In this work, existing a literature suggests a "preferred" distribution for σ_0 (Hess et al. [1] suggest a lognormal distribution). According to this informative knowledge, a prior model probability $\pi_{LN} = 0.9$ is assigned and equal probability ($\pi_j = \frac{1-0.9}{6},\ j \neq LN$) for the other models is assumed for our problem. We refer to this as the "strong correct" prior since a strong belief exists in the correct prior model. This strong correct prior will be compared against the uniform prior of equal probabilities and a "strong incorrect" prior wherein there is a strong belief in the incorrect Log-logistic model such that it has prior probability $\pi_{LL} = 0.9$ and all other prior

probabilities are equal. Table 3.2 summarizes the three model prior cases.

Table 3.2: Prior model probabilities.

	Uniform	"Strong Correct"	"Strong Incorrect"
Gamma	1/7	0.0167	0.0167
Inverse Gaussian	1/7	0.0167	0.0167
Logistic	1/7	0.0167	0.0167
Log-logistic	1/7	0.0167	0.9
Lognormal	1/7	0.9	0.0167
Normal	1/7	0.0167	0.0167
Weibull	1/7	0.0167	0.0167

Posterior model probabilities are updated according to Eqs. (2.11)-(2.12) as data are added into this model. These probabilities rely on the assumed model parameter prior and consequently they differ based on the historical data that we use to construct the prior. Note that, when data is very limited, it is very difficult to make any meaningful conclusions with respect to the model probabilities as evidenced by the data in Table 3.3, which gives the posterior model probabilities from 10 yield stress data for each of the parameter priors given equal prior model probabilities. In fact, the posterior is simply equal to the model evidence in these cases. In short, this is a classic small data case where a precise "best" model is impossible to identify. Furthermore, these posterior model probabilities are dependent on the parameter prior with significant differences across different priors as suggested by the definition of model evidence.

As a function of the number of collected data, the convergence of the model-form uncertainty is also of interest. In Table 3.3, we highlighted the rela-

Table 3.3: Posterior model probabilities given initial 10 data and different parameter priors given equal model prior probabilities.

Distribution	AIC	Noninformative	ABS-A	ABS-B	ABS-C	ASTM-A7
Gamma	0.168	0.167	0.159	0.157	0.170	0.166
Inverse Gaussian	0.172	0.184	0.142	0.150	0.132	0.191
Logistic	0.119	0.115	0.161	0.118	0.064	0.136
Loglogistic	0.128	0.125	0.182	0.096	0.063	0.163
Lognormal	0.167	0.162	0.184	0.140	0.182	0.176
Normal	0.154	0.149	0.147	0.178	0.189	0.130
Weibull	0.091	0.098	0.024	0.160	0.201	0.037

tionship between the very small datasets and large model-form uncertainties – with further uncertainty caused by the choice of the parameter prior. How then does the performance change with various parameter priors and how much data is necessary to reduce the uncertainty to acceptable level?

Fig. 3.4, the posterior model probabilities are presented as a function of dataset size for different parameter priors when the prior model probabilities are equal. For comparison, the posterior model probabilities using AIC model selection (i.e. given savvy prior probabilities) are shown in Fig. 3.5. The non-informative, ABS-B and AIC priors show almost identical trends as data are collected. All of them essentially identify the Inverse Gaussian model and Lognormal model with equal probability and fail to identify a unique model. However, they are among the "best" priors given that the Lognormal and Inverse Gaussian models are nearly identical in this case. Also noteworthy is that the ABS-A parameter prior converges toward the wrong Gamma model and effectively discounts the Lognormal model entirely.

The previous discussion suggests that informative model priors can signif-

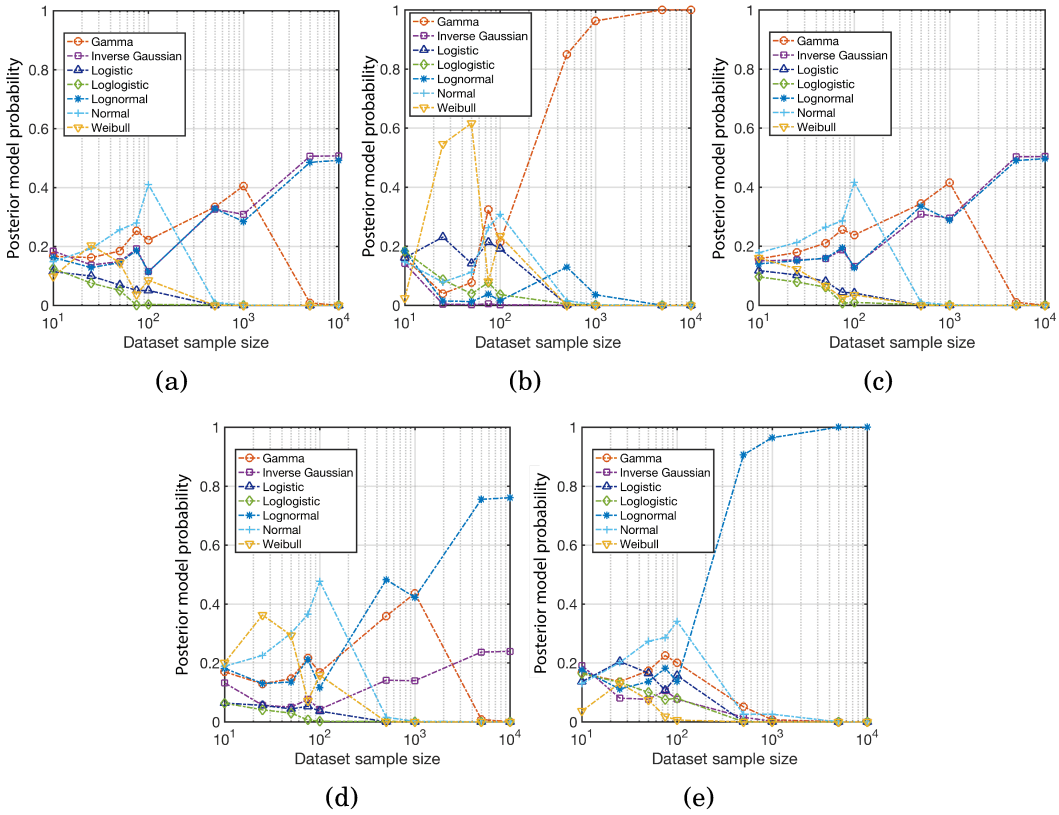


Figure 3.4: Posterior model probabilities given equal prior model probabilities as a function of dataset size for different parameter priors: (a) Noninformative prior (b) ABS-A prior (c) ABS-B prior (d) ABS-C prior (e) ASTM-A7 prior

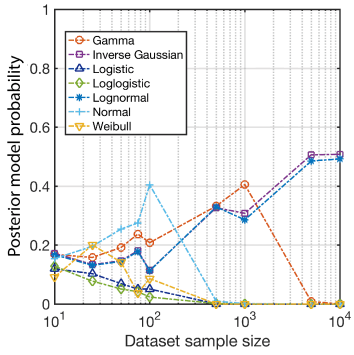


Figure 3.5: Posterior model probabilities from AIC model selection.

icantly change the convergence behavior as data are collected. For each of the seven models, Fig. 3.6 and Fig. 3.7 illustrate the posterior model proba-

bilities with dataset size based on the strong correct prior (Fig. 3.6) and the strong incorrect prior (Fig. 3.7) for each of parameter priors. In the case of the strong correct model prior, four of the five cases show convergence toward the true Lognormal model when the dataset grows large. Even in the case of the strong incorrect model prior probabilities, the Bayesian multimodel methodology eventually suppresses the incorrect Log-logistic model and identifies the correct Lognormal model in these cases. This is to say, there is a degree of robustness for these parameter priors. Notice that the ABS-B parameter prior with strong incorrect model prior leads to equal posterior model probability for the Inverse Gaussian and Lognormal models. Again, this is because these two distribution models are almost identical in shape such that it is not easy to discern between them in the inference given the prior information. In both cases, the ABS-A parameter prior leads to the selection of the incorrect Gamma model even 10,000 data are collected. In other words, it may be impossible to infer even the correct model for the data if the parameter prior is not wisely selected. In practice, this conclusion is critical for uncertainty quantification and propagation.

3.2.2.2 Effect of parameter prior on parameter uncertainty

The parameter prior will play a very important role in the convergence of the parameter posterior for each probability model form. In this work, we select

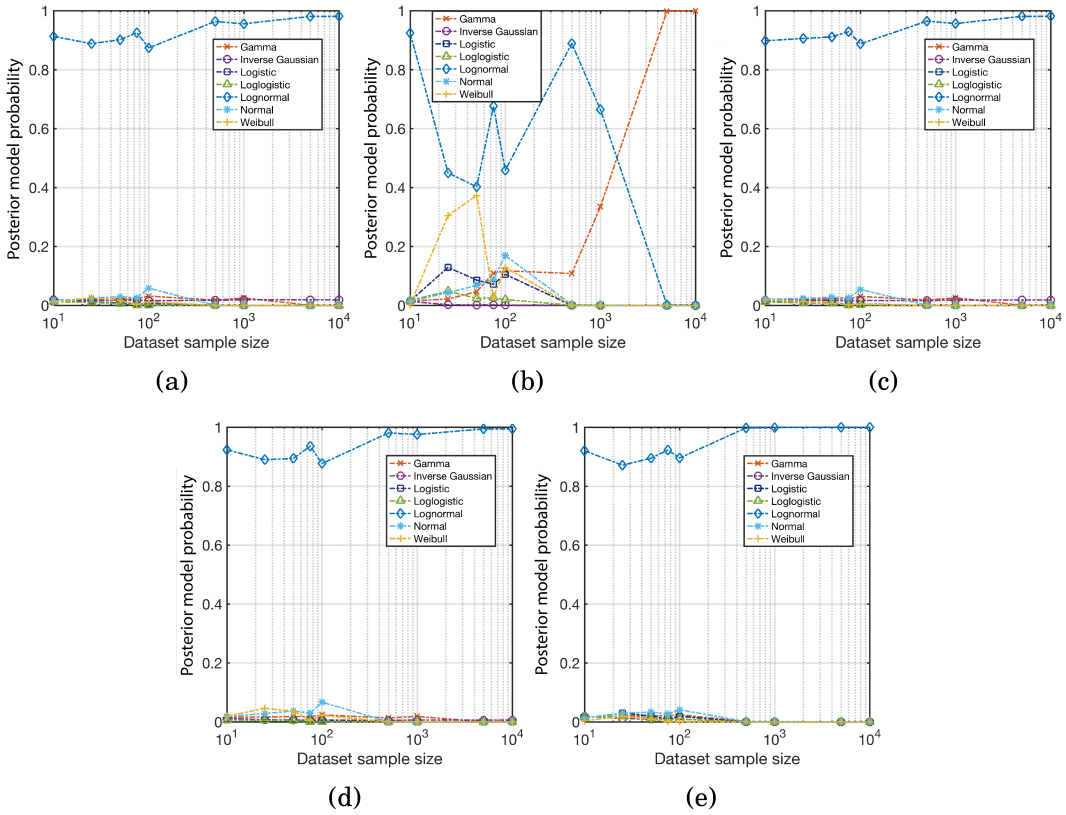


Figure 3.6: Posterior model probabilities given "strong correct" prior model probabilities as a function of dataset size for different parameter priors: (a) Noninformative prior (b) ABS-A prior (c) ABS-B prior (d) ABS-C prior (e) ASTM-A7 prior.

the Lognormal distribution as a representative case to investigate the effect. Similar performance of the other models was observed.

We draw data from the "true" Lognormal distribution and conduct Bayesian inference to estimate the parameters of the Lognormal distribution using each of the five considered parameter priors. Table 3.4 shows the posterior joint parameter densities given "small" datasets ($\leq$ 100 data) along with the true parameters (indicated by a $\star$).

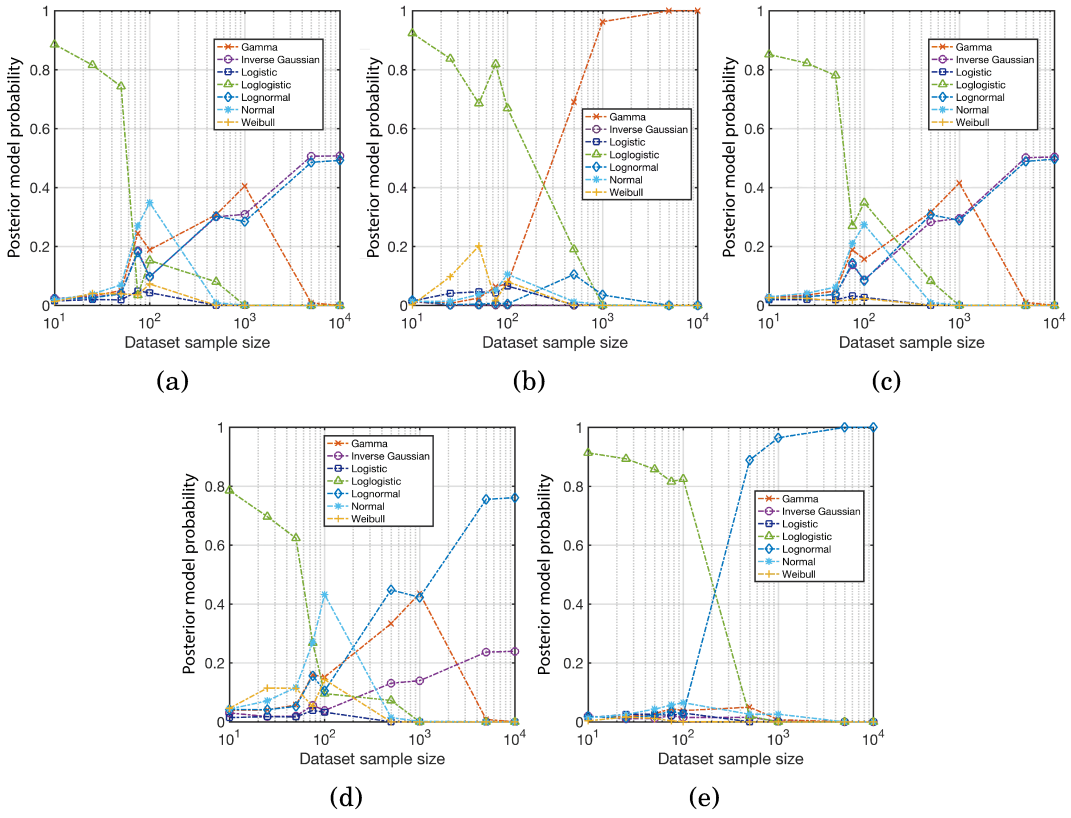
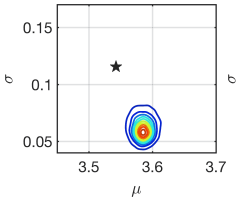
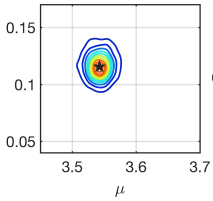
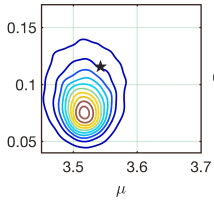
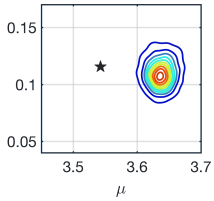
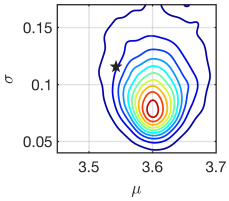
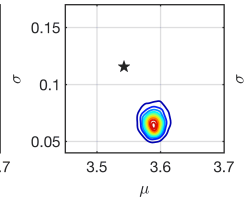
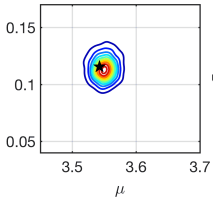
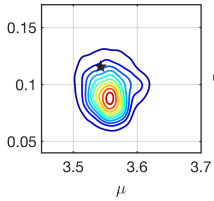
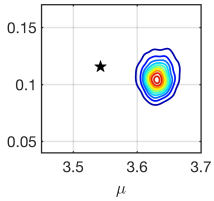
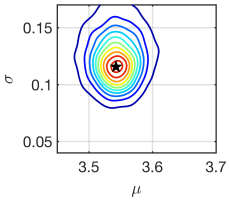
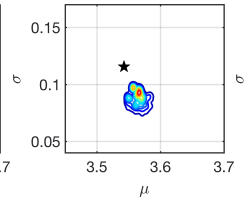
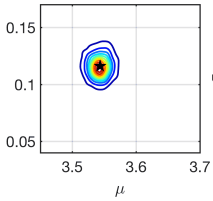
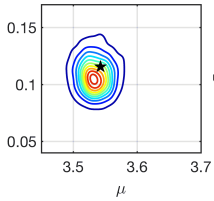
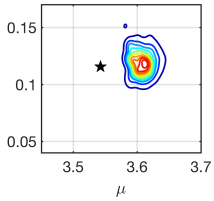
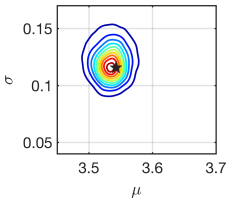
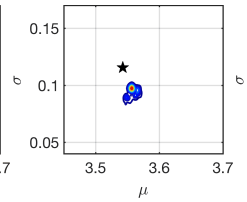
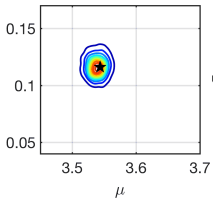
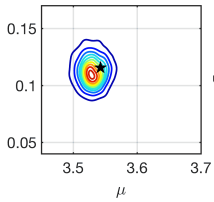
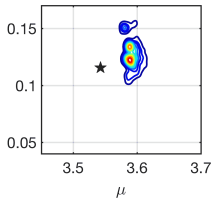
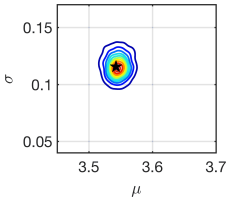
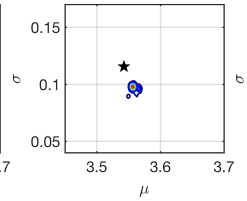
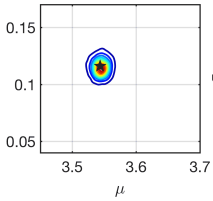
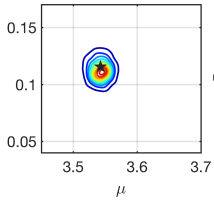
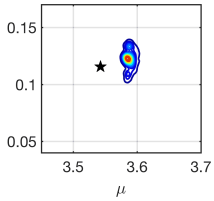


Figure 3.7: Posterior model probabilities given "strong incorrect" prior model probabilities as a function of dataset size for different parameter priors: (a) Noninformative prior (b) ABS-A prior (c) ABS-B prior (d) ABS-C prior (e) ASTM-A7 prior.

Note that, even though ABS-A and ASTM-A7 priors converges very rapidly, neither their priors not their posteriors include the true parameter values. From small datasets, these models cannot infer the correct parameter distribution. The ABS-C and noninformative parameter priors show relatively similar rates of convergence and the posteriors include the true estimator. The principal reason is that the amount of incorrect information in both of them are sufficiently weak. Finally, the correct ABS-B prior shows the best convergence

Table 3.4: Posterior parameter joint probability densities for the lognormal distribution with different priors considering small dataset size (≤ 100 data).

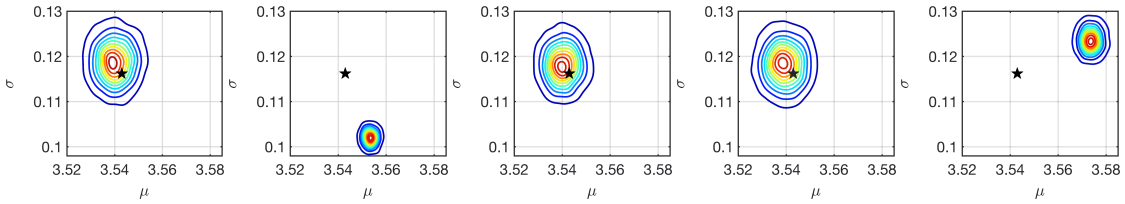
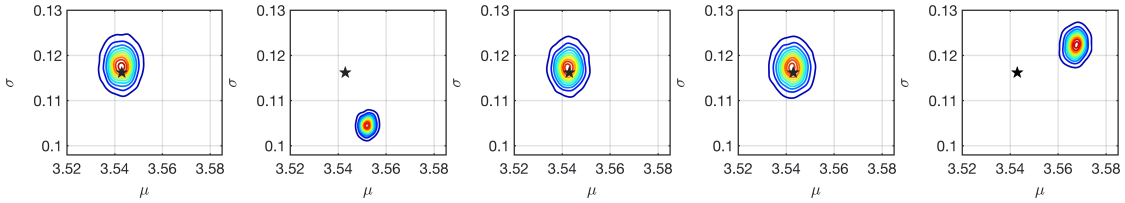
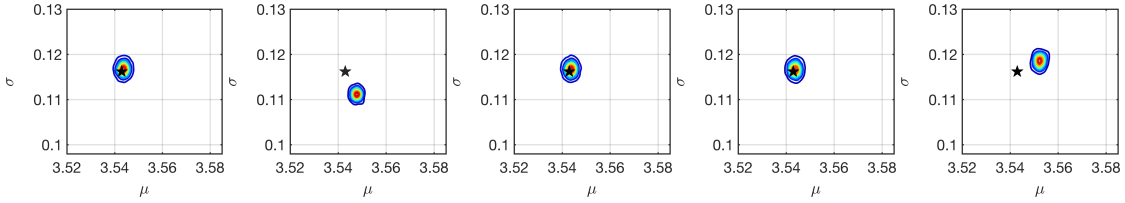
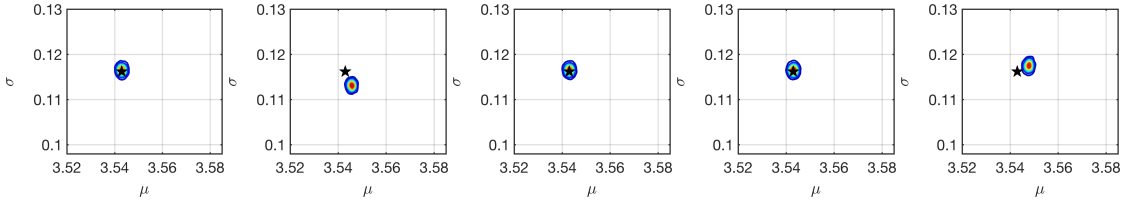
Data	Noninformative	ABS-A (33)	ABS-B (79)	ABS-C (13)	ASTM-A7 (58)
Prior					
10					
25					
50					
100					

to the true model as expected.

A set of similar plots for "large" datasets ($\geq$ 500 data) are shown in Table 3.5. Notice that the ABS-A and ASTM-A7 priors narrow continuously and move slowly toward to the correct estimator of model parameters. But they do not include the correct parameters in the posterior joint densities even after 10,000

data are collected. On the contrary, the noninformative, ABS-B and ABS-C priors continue to converge correctly at similar rates of convergence.

Table 3.5: Posterior parameter joint probability densities for the lognormal distribution with different priors considering large dataset size (≥ 500 data).

Data	Noninformative	ABS-A (33)	ABS-B (79)	ABS-C (13)	ASTM-A7 (58)
500					
1000					
5000					
10000					

As an alternative way of viewing this effect, we populate a number of possible distributions using Monte Carlo sampling from the posterior joint parameter distributions. We can see how the distributions change with the dataset size for the noninformative, ABS-B, and ABS-A parameter priors in Table 3.6. We see that the true distribution is included by the band of distributions for the noninformative and ABS-B priors but is not for the ABS-A prior. However,

the band of distributions from the ABS-A prior is much narrower than those from the noninformative and ABS-B priors. In other words, the ABS-A prior places a high degree of confidence in incorrect distributions. This may have major implications for uncertainty propagation. .

3.2.2.3 Effect of priors on total uncertainty

In Chapter 2, the total uncertainty is represented by Monte Carlo sampling from the candidate models. To obtain each sample (pdf), a probability model is randomly selected based on the posterior model probabilities, and the model parameters are then randomly selected from its posterior joint densities. Fig. 3.8 shows the Monte Carlo set of distributions for different dataset size given equal model prior probabilities and noninformative parameter priors.

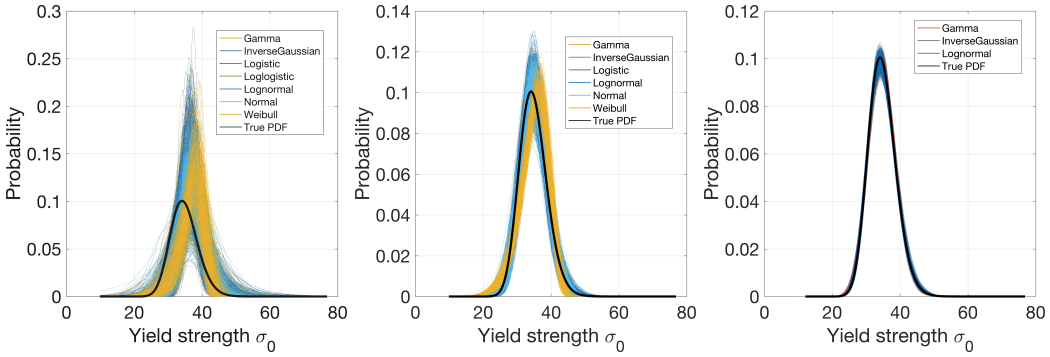
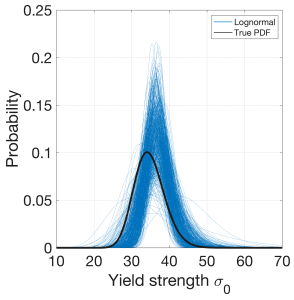
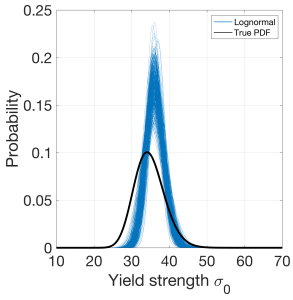
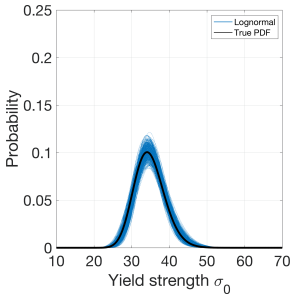
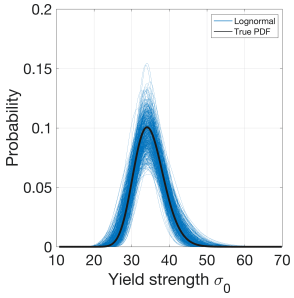
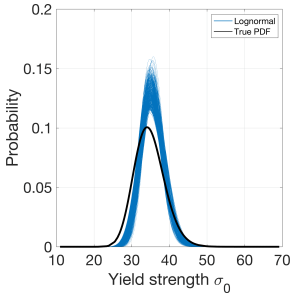
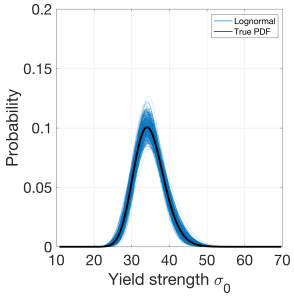
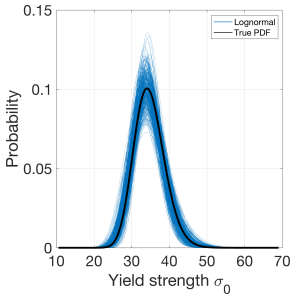
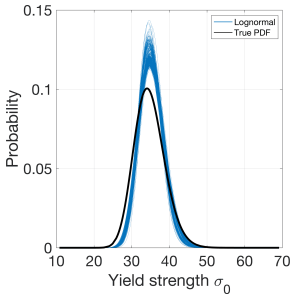
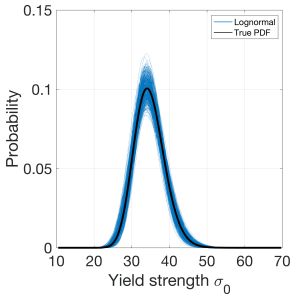
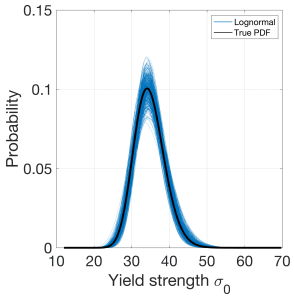
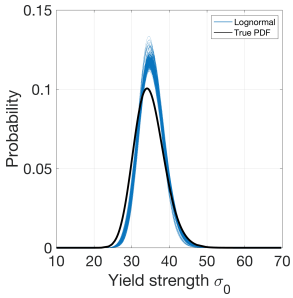
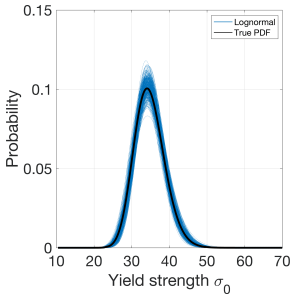


Figure 3.8: 5000 distributions given equal model prior probabilities with noninformative parameter priors for (a) 10data, (b) 100 data and (c) 1000 data

To observe the uncertainty level in a specified model set, we define a metric that is referred to as the average mean square distance between the 5000

Table 3.6: Monte Carlo sets of lognormal distributions drawn from the posterior parameter densities given noninformative, ABS-A, and ABS-B prior parameter densities.

Data	Noninformative	ABS-A (33)	ABS-B (79)
10			
25			
50			
100			

models in the set and the true Lognormal distribution given by:

$$\delta = \frac{1}{2}\frac{1}{5000}\sum_{i=1}^{5000}(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - p(\boldsymbol{x}))^2 \tag{3.12}$$

where $p(\mathrm{x})$ is the Lognormal distribution of the true model and $p_i(\mathrm{x}|\boldsymbol{\theta})$ are the distributions in the set. Fig. 3.9 shows the distance as a function of dataset size for each parameter prior and for the three cases of model prior probabilities.

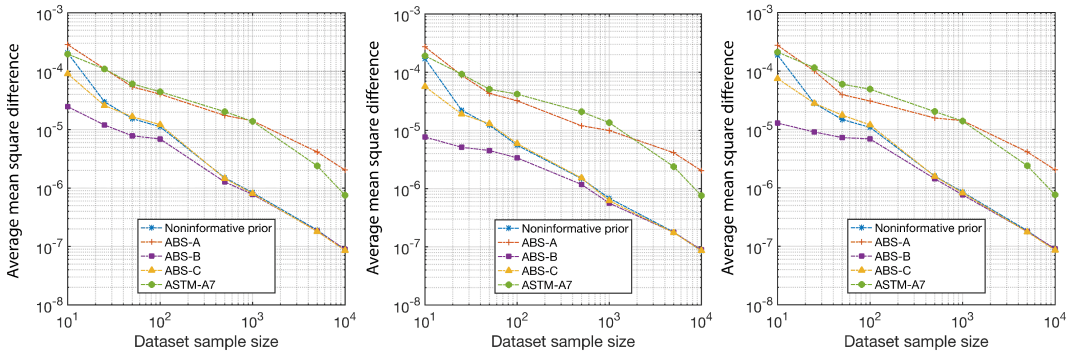


Figure 3.9: Convergence of average mean square distance for (a) equal model prior, (b) strong correct model prior and (c) strong incorrect model prior

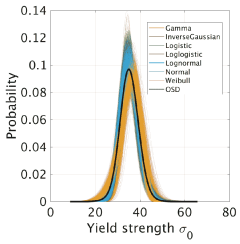
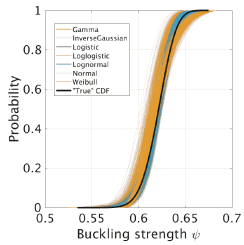
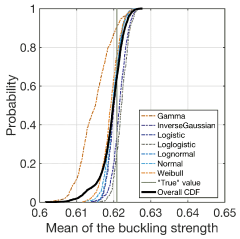
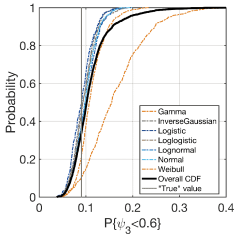
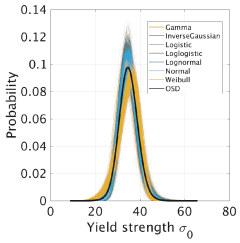
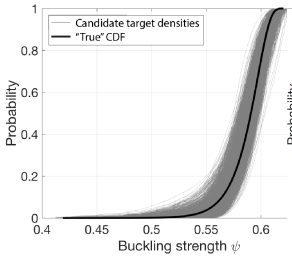
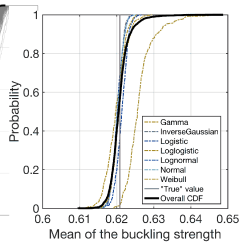
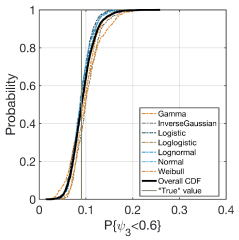
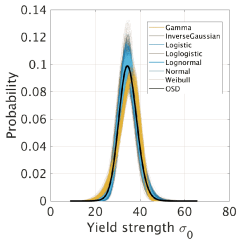
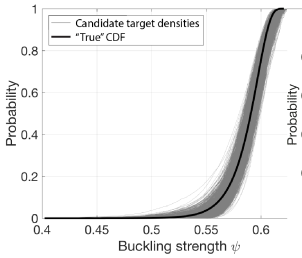
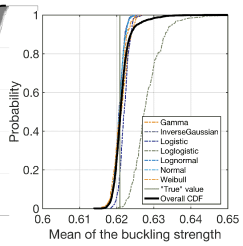
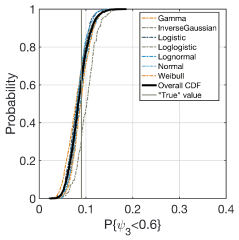
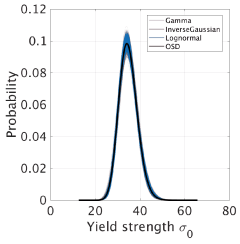
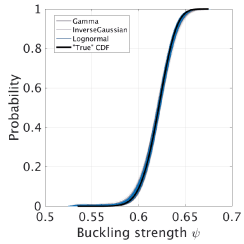
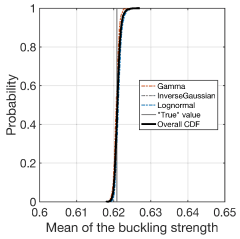
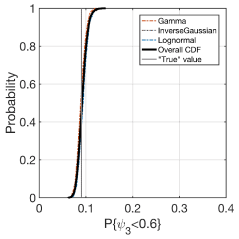
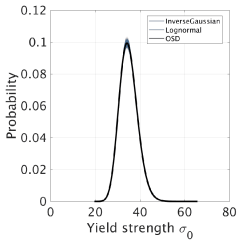
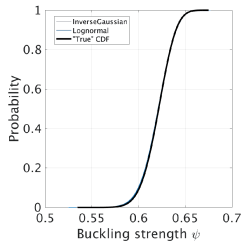
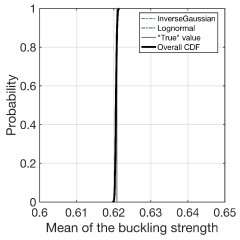
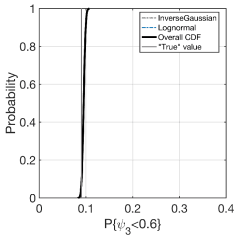
In the small data case, using the correct ABS-B prior brings a major benefit. The set of posterior distributions compares favorably to the true distribution. In the meantime, all other priors poorly represent the true distribution. After 100 data are collected, both the noninformative and ABS-C priors are nearly as good as the ABS-B prior. On the other hand, the ABS-A and ASTM-A7 priors do not yield similar accuracy as the other priors even when large datasets are available. We may therefore conclude that the set of distributions drawn from the ABS-A and ASTM-A7 priors have residual errors even for large datasets, which effectively yields identification/propagation of incorrect probability models. This is reflected again by the Lognormal distributions from the ABS-A prior in Table 3.6 that do not include the true model.

3.2.3 Influence of data-driven priors on uncertainty propagation

A number of probability models are identified by the Bayesian multimodel methodology (e.g. Figure 3.8). These models are then propagated through a computational model according to the method [3] presented in Chapter 2. Considering the sensitivity of the posterior probabilities to the choice of the priors, it will raise a question: What influence do the prior assumptions have on output quantities of interest from the model? If the convergence of prior is rapid, then we should also expect rapid convergence in output quantities of interest. However, if a bad prior is chosen, how poor are the results? Or if a good prior is selected, how much of an improvement can be obtained? As discussed in the previous section, the results seem to imply that a poor prior not only yields incorrect probabilistic response but also causes large uncertainties. In this section, these issues will be fully explored in terms of the plate buckling strength problem.

Let us focus on the ABS-B parameter prior with equal prior model probabilities for illustration. Table 3.7 shows the uncertainty propagation results for various dataset size. The left column in Table 3.7 is referred to as the set of 5000 probability models identified from the Monte Carlo sampling of the quantified uncertainties in model parameter and model-form. The bold curves show

Table 3.7: Optimal sampling density (OSD), CDFs, mean and probability of failure for ABS-B prior associated with equal model prior probability as a function of dataset size from 10, 25, 50, 500, to 5000

Data	OSD	CDFs	Mean	Probability of failure
10				
25				
50				
500				
5000				

the optimal sampling density that is used for the propagation of uncertainties.

The second column in Table 3.7 shows the CDFs for the buckling strength along

with the true CDF given the Lognormal model. We can see that the true CDF is fully included within the set of propagated distributions. The third and fourth column in Table 3.7 show the CDFs of the mean buckling strength and probability of failure ($P_f = P(\psi < 0.6)$). The colored CDFs are conditional CDFs for each probability model form while the bold black curve presents the overall CDF considering all model forms with their probabilities. In all cases, the true mean buckling strength and the probability of failure all fall within the range of the CDFs.

With increasing dataset size, the uncertainty diminishes as expected. This can be noted that the band of distributions in both input PDFs and output CDFs gradually narrow toward their correct distributions (true estimator). Similarly, the range of the CDFs for the mean buckling strength and the probability of failure gradually narrow toward the true values when more data are collected.

When a good prior is selected, all of these trends show the method's performance. However, are the same trends observed for other priors? To investigate, we propose two different metrics to quantify the convergence of the mean buckling strength and variance of buckling strength under different priors. One is a simple quantile confidence metric that defines the 95% confidence range for

statistic Y given n data by:

$$\delta_Y^{(n)} = Q_{0.975}(Y^{(n)}) - Q_{0.025}(Y^{(n)}) \tag{3.13}$$

The ranges for the mean, variance, and P_f are therefore denoted $\delta_\mu^{(n)}$, $\delta_{\sigma^2}^{(n)}$, and $\delta_{(\psi<\psi^*)}^{(n)}$. The other one is a relative accuracy metric, the "area validation metric" [134,135], which measures the difference in area between the CDF and the true value for statistic Y given n data as:

$$d_Y^{(n)}(F,T) = \int_{-\infty}^{\infty} |F(Y) - T(Y)|\, dy \tag{3.14}$$

Where $T(Y)$ is the true value and $F(Y)$ is the CDF from the simulation. For the mean, variance, and P_f the accuracy metrics are therefore denoted by $d_\mu^{(n)}$, $d_{\sigma^2}^{(n)}$, and $d_{(\psi<\psi^*)}^{(n)}$, respectively.

Given the correct ABS-B parameter prior, we first study the effect of the prior model probability. The convergence of the confidence metric (Eq. (3.13)) and area accuracy metric (Eq. (3.14)) are shown in Fig. 3.10 respectively. From these figures, we conclude that the strong correct model prior probabilities significantly improve the confidence and accuracy for the mean and standard deviation of the buckling strength when only small datasets are available. But the improvement will diminish as more data are collected. Meanwhile, the prior probabilities have a relatively modest effect on the convergence of the

probability of failure.

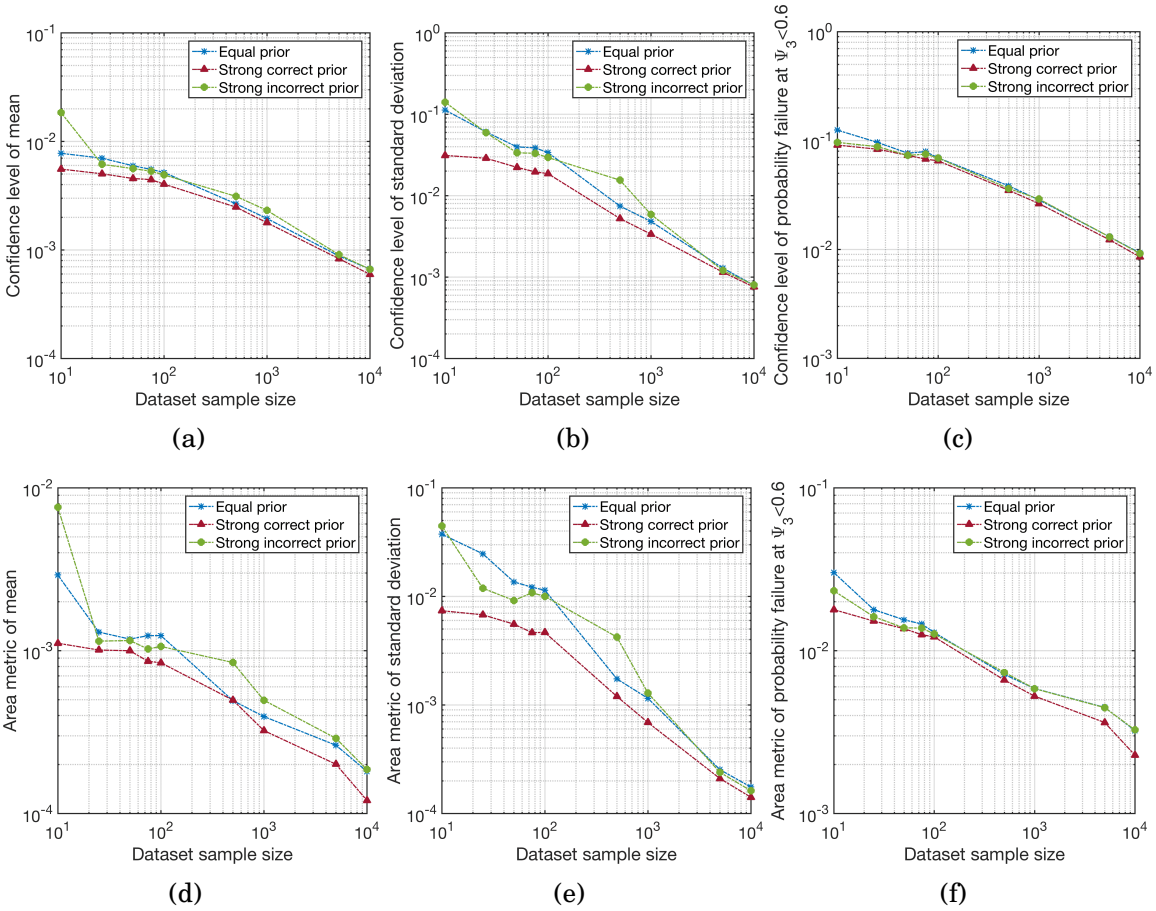


Figure 3.10: Compare the effect of prior model probability for ABS-B prior - convergence of confidence level of (a) mean, (b) variance (c) probability of failure; and area validation metric for (d) mean, (e) variance and (f) probability of failure

Next, we investigate the effect of the parameter priors. Given equal prior model probabilities, the parameter prior is varying. Fig. 3.11 shows the convergence of the confidence and area metrics for the mean buckling strength, variance of buckling strength, and P_f with dataset size. ABS-B prior, as a good prior, shows good performance on both confidence and accuracy as expected.

Most of the other priors also present reasonable performance and converge at approximately the same rate. The main problem appears in the accuracy convergence of mean buckling strength and probability of failure based on the ASB-A and ASTM-A7 priors. Recall that these models did not accurately quantify the input uncertainty. Thus, these priors fail to show convergence in terms of the accuracy of response statistics.

But Fig. 3.11 (a)-(c) shows consistent convergences in confidence level regardless of the prior, while Fig. 3.11 (d)-(f) suggests that accuracy depends strongly the prior. Therefore, cases with poor parameter prior exhibit high confidence in inaccurate statistics. Fig. 3.12, as a more clear illustration, shows the CDFs for the mean buckling strength and the probability of failure for different priors given 10,000 data with equal prior model probabilities. The CDF of the mean value for the ASTM-A7 prior does not intersect the true value but has a similar spread as the other distributions. From the quantitative perspective, its confidence metric is small and shows that 95% probability lies in the range $[0.61979, 0.62036]$ but inaccurate given the true value is $\mu_\psi = 0.62089$. The ABS-A and ASTM-A7 priors also yield high confidence in incorrect estimates of the probability of failure as shown in Fig. 3.12 (b). Their values of 95% probability lie in the range $[0.09758, 0.10725]$ and $[0.07249, 0.07974]$ respectively but both are inaccurate given the true value is $P_f = 0.090132$. In order words, even for very large dataset size, using these priors gives high confidence in inaccu-

rate estimators.

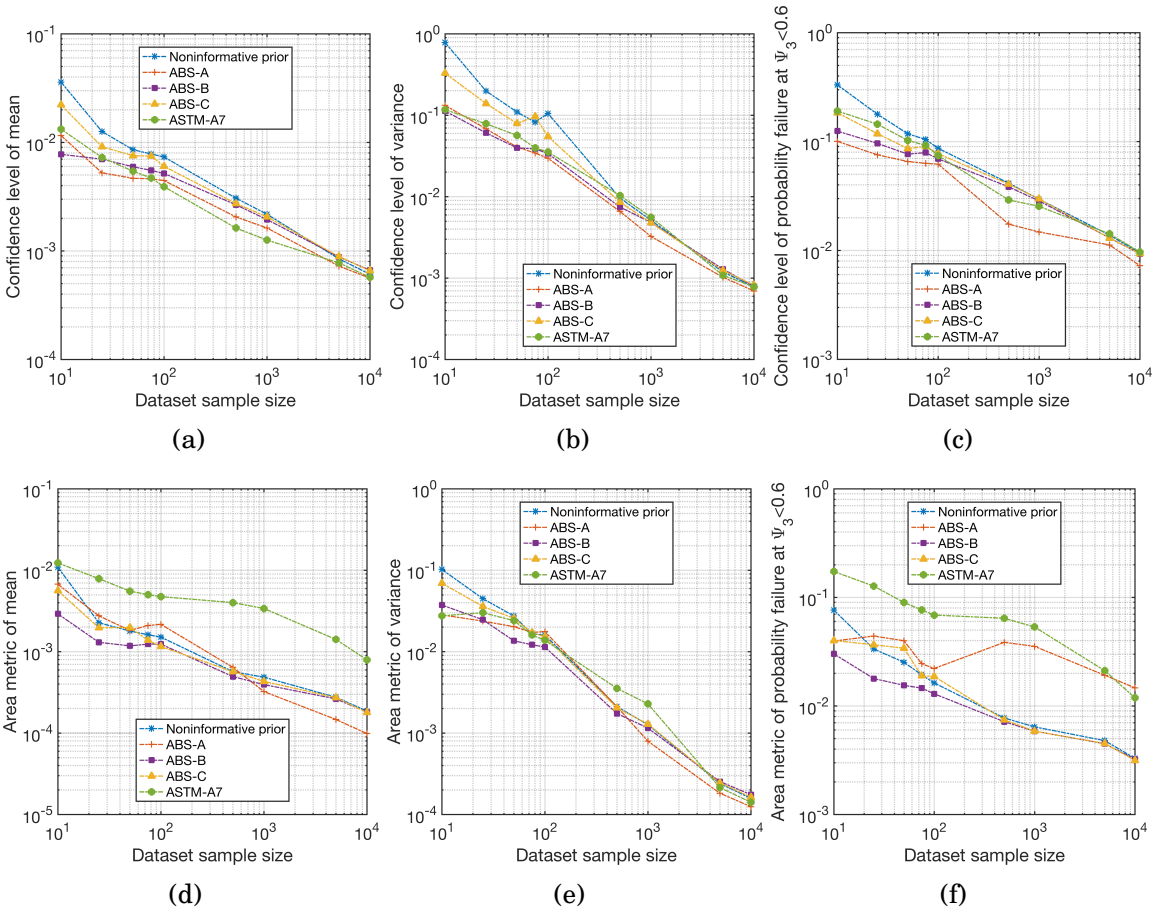
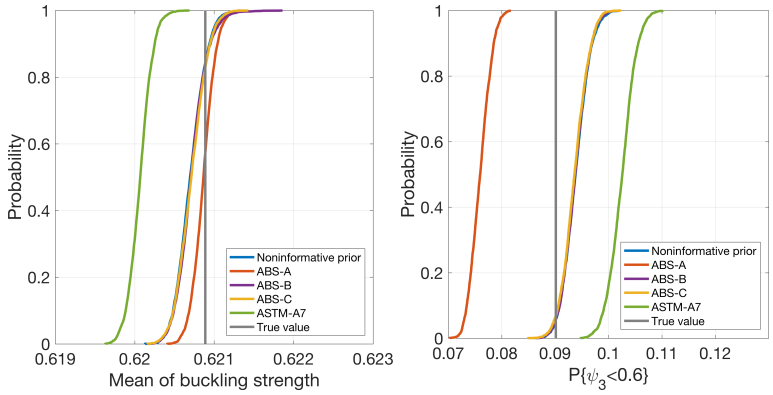


Figure 3.11: Equal prior model probability and different parameter priors - convergence of confidence level of (a) mean, (b) variance and (c) probability of failure; and area validation metric for (d) mean, (e) variance and (f) probability of failure

3.3 Conclusion

This chapter primarily focuses on understanding the effect of prior probabilities in both probability model-form and model parameters on multimodel



(a) (b)

Figure 3.12: Empirical CDFs of (a) mean of buckling strength and (b) probability of failure at $\psi_3 < 0.6$ given 10000 data with equal prior model probability for different parameter priors

uncertainty quantification and propagation within a fully Bayesian framework. Through an example considering plate buckling strength problem, we systematically explore the effect of various prior model-form and model parameter probabilities on quantification and propagation of uncertainties resulting from small datasets. In terms of model-form uncertainties, the assumptions about prior probabilities show a significant impact on the quantified uncertainties given small data but incorrect prior probabilities can be overcome by large datasets if the parameter priors are reasonable. Furthermore, parameter priors derived from the historical datasets that are similar to the presently collected data (but nonetheless different) can introduce biases in the multimodel inference that persist even as very large datasets are collected. The combined effects of model-form and model parameter priors on uncertainty propagation are then investigated. Again, it is shown that uncertainties in response quan-

tities depend strongly on both priors and biases introduced by incorrect priors persist yielding inaccurate probabilistic response quantities even in the large data limit.

Chapter 4

Uncertainty quantification and propagation with dependence modeling

In engineering applications, it is common to assume that the model inputs are mutually independent or modeled by a Gaussian dependence structure. A number of UQ approaches require a transformation of the model inputs into independent variables. However, this transformation is difficult to compute in many cases when considering the Gaussian correlation modeling, which admits an analytical solution (Nataf transformation [136, 137]). For this reason, the Gaussian correlation assumption is widely applied in the context of UQ even though it may be inaccurate, particular when available data are very lim-

ited. This chapter aims to address this challenge by modeling the dependence structure of multivariate inputs using copula theory. Using the multimodel approach developed herein, we investigate the uncertainty resulting from small datasets in both model dependence structure and marginals. The overall uncertainties are then illustrated for a composite material problem.

4.1 Copula-based modeling of dependence structure

4.1.1 Dependency measures

The most well known measure of dependence between quantities is the Pearson's correlation coefficient, commonly named simply the correlation coefficient, but it only measures linear dependence. Considering two random variables X and Y with mean values μ_X and μ_Y and standard deviation σ_X and σ_Y, the correlation coefficient $\rho_{X,Y}$ is defined as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{4.1}$$

where $E[\cdot]$ is the expected value operator and cov is the covariance. All correlation coefficient values are bounded in the interval $[-1, 1]$, indicating the degree

of linear dependence between two variables. The closer the coefficient is to either 1 or -1, the stronger the correlation between the variables. If the variables are independent, the correlation coefficient is 0.

Another common measure of dependence is Kendall's τ, which measures the difference between the concordance and discordance probability and can be used to detect the nonlinear dependence. Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed random variables, then Kendall's tau is defined as

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{4.2}$$

However, the information given by a correlation coefficient (Pearson's ρ or Kendall's τ) is not enough to define the dependence structure between random variables (except in special cases, e.g. Gaussian random variables). One method to capture a more complete view of dependence structure is to model the dependence using a copula. In practice, many data structures exhibit different marginal distributions, nonsymmetric dependencies as well as heavy tail dependencies between some pairs of variables. These variables cannot be modeled by a Gaussian or multivariate t distribution. This challenge is overcome by the copula approach that allows to model dependencies and marginal distributions separately.

4.1.2 Copula theory

Consider F as the d-dimensional distribution function of the random vector $\boldsymbol{X} = (X_1, ..., X_d)^T$ with univariate marginal distributions $F_1, ..., F_d$. According to Sklar's theorem [138], there exists a copula C such that for all $\boldsymbol{x} = (x_1, ..., x_d)^T \in [-\infty, \infty]^d$,

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C_{1,...,d}(F_1(x_1), ..., F_k(x_d)) \tag{4.3}$$

If $F_1, ..., F_d$ are continuous, the copula C is unique. The copula C can be interpreted as the distribution function of a d-dimensional random vector on $[0, 1]^d$ with uniform univariate marginals.

Sklar's theorem can also be restated with respect to probability densities. The corresponding copula density can be expressed as:

$$c_{1,...,d}(F_1(x_1), ..., F_d(x_d)) = \frac{\partial C_{1,...,d}(F_1(x_1), ..., F_d(x_d))}{\partial F_1(x_1), .., \partial F_d(x_d)} \tag{4.4}$$

which implies the joint multivariate pdf can be formulated by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = c_{1,...,d}(F_1(x_1), ..., F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d) \tag{4.5}$$

where $f_k(x_k), 1 \leq k \leq d$ is the marginal pdf. For the bivariate case, Joe [70] and Nelsen [139] provided a rich variety of copula families from the two major classes of *Elliptical* and *Archimedean* copulas. Elliptical copulas are directly

derived by inverting Sklar's theorem, shown in Eq. (4.3). Given a bivariate cumulative distribution function F with marginals F_1 and F_2, then

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \tag{4.6}$$

is a bivariate copula for $u_1, u_2 \in [0, 1]$. One of the most commonly used bivariate elliptical copula is the bivariate Gaussian copula

$$C(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{4.7}$$

where Φ_ρ is the joint cumulative distribution of bivariate standard normal distribution with correlation parameter ρ and Φ^{-1} is the inverse standard normal cumulative distribution function.

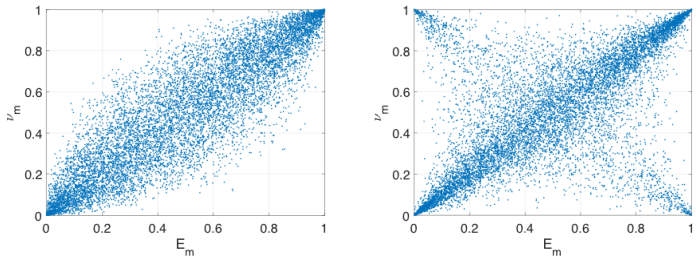
Another common copula is Student-t copula. The bivariate Student-t density is given by

$$f(\boldsymbol{x}) = \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^2|\Sigma|}} \left(1 + \frac{(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+2}{2}} \tag{4.8}$$

where ν is the number of degrees of freedom, $\boldsymbol{\mu}$ is the mean vector and Σ is a positive-definite matrix. The corresponding Student-t copula is given by

$$C(u_1, u_2) = t_{p,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \tag{4.9}$$

where t_ν^{-1} is defined as the inverse univariate Student-t marginal distribution function with ν degrees of freedom and $t_{p,\nu}$ is the bivariate Student-t cumulative distribution function with correlation parameter p implied by the matrix Σ. Fig. 4.1 shows samples from the elliptical copula family with Gaussian copula and Student-t copula. Table 4.1 provides the basic properties of the Gaussian copula and the Student-t copula.



(a) (b)

Figure 4.1: Elliptical copula family (a) Gaussian copula and (b) Student-t copula

Table 4.1: Properties and definition of elliptical copula families

Elliptical family	Parameter range	Kendall's τ	Tail dependence
Gaussian	$\rho \in (-1, 1)$	$\frac{2}{\pi} arcsin(\rho)$	0
Student-t	$\rho \in (-1, 1), \nu > 2$	$\frac{2}{\pi} arcsin(\rho)$	$2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}})$

Another important copula family, Archimedean copulas are defined as

$$C(u_1, u_2) = \psi^{[-1]}(\psi(u_1) + \psi(u_2)) \tag{4.10}$$

where ψ is the generator function of the copula C, which is a continuous strictly

decreasing convex function which satisfies $\psi(1) = 0$ and $\psi^{[-1]}$ is defined as

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t), & 0 \leq t \leq \psi(0) \\ 0, & \psi(0) \leq t \leq \infty \end{cases} \tag{4.11}$$

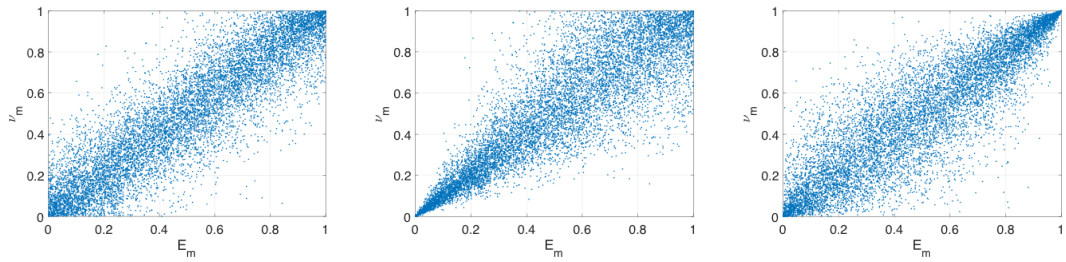
The most common single parameter Archimedean copulas are the Clayton, Gumbel and Frank [139]. Their bivariate copula formulations are shown in Table 4.2, with their corresponding properties (generator and Kendall's τ) shown in Table 4.3 where $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$ is Debye function [70, 139]. Fig. 4.2 show examples of samples drawn from these copulas for two random variables E_m and ν_m.

Table 4.2: Definitions of Archimedean copula families

Name of Copula	Bivariate copula $C_\theta(u_1, u_2)$	Parameter θ
Clayton	$\left[\max\left\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\right\}\right]^{-1/\theta}$	$\theta \in [-1, \infty) \setminus \{0\}$
Frank	$-\frac{1}{\theta}\log\left[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$	$\theta \in \mathbb{R} \setminus \{0\}$
Gumbel	$e^{-\left((-\log(u_1))^\theta + (-\log(u_2))^\theta\right)^{1/\theta}}$	$\theta \in [1, \infty)$

Table 4.3: Properties of Archimedean copula families

Name of Copula	Generator	Kendall's τ
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\frac{\theta}{\theta+2}$
Frank	$-\log[\frac{e^{-\theta t}-1}{e^{-\theta}-1}]$	$1 - \frac{4}{\theta} + 4\frac{D_1(\theta)}{\theta}$
Gumbel	$(-\log t)^\theta$	$1 - \frac{1}{\theta}$



(a) (b) (c)

Figure 4.2: Archimedean copula family (a) Frank copula, (b) Clayton copula and (c) Gumbel copula

4.1.3 Vine copulas

Copula families perform well in the bivariate case, but in arbitrary dimension, particularly high dimension, the choice of adequate copula families is very limited. Elliptical families and Archimedean copulas lack the flexibility to accurately model the dependence structure of high dimensional variables. Simple extension of these bivariate families offer some improvement, but typically become intricate and introduce additional limitations that can not be applied to establish a distribution consistent with arbitrary correlation [140].

Vine copulas (also called tree structures) do not suffer from these issues and have been widely used in many fields of application. Bedford and Cooke [69] introduced a graphical model for describing multivariate copulas using a cascade of bivariate copulas, denoted by *pair-copulas*. This pair-copula construction provides a flexible way to decompose a multivariate probability density into bivariate copulas such that each pair-copula is independent of the others.

Consider an n-dimensional joint density function $f(x_1, ..., x_n)$ for a random vector $\boldsymbol{X} = (X_1, ..., X_n)$. This density can be decomposed based on the law of total probability

$$f(x_1, ..., x_n) = f_n(x_n) \cdot f(x_{n-1}|x_n) \cdot f(x_{n-2}|x_{n-1}, x_n) \cdots f(x_1|x_2, ..., x_n) \tag{4.12}$$

From Sklar's theorem, we also know the joint probability density can be formulated as shown in Eq. (4.5). In the bivariate case, Eq. (4.5) simplifies to

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \tag{4.13}$$

where c_{12} is the appropriate *pair-copula density* for the pair of transformed variables $F_1(x_1)$ and $F_2(x_2)$. It is straightforward to write a conditional density

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \tag{4.14}$$

in terms of the pair-copula. Similarly, it easily follows for three random variables X_1, X_2 and X_3 as follows

$$f(x_1|x_2, x_3) = c_{12|3}(F(x_1|x_3), F(x_2|x_3)) \cdot f(x_1|x_3) \tag{4.15}$$

for the appropriate pair-copula $c_{12|3}$ which is used for he transformed variables

$F(x_1|x_3)$ and $F(x_2|x_3)$. An alternative decomposition is

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot f(x_1|x_2) \tag{4.16}$$

where $c_{13|2}$ differs from the pari-copula in Eq. (4.15). We can further decompose $f(x_1|x_2)$ in Eq. (4.16) based on Eq. (4.14)

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \tag{4.17}$$

By the explanation in the above cases, it is now clear that the conditional marginal term in Eq. (4.5) can be decomposed into the appropriate pair-copula using the general form given by [141, 142]

$$f(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j}))f(x|\boldsymbol{v}_{-j}) \tag{4.18}$$

where v_j is an arbitrarily excluded element from vector $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the vector $\boldsymbol{v}$ after excluding v_j. Hence, a multivariate density $f(\boldsymbol{x})$ can be expressed as a product of bivariate copula density functions with marginal conditional CDFs in the form of $F(\boldsymbol{x}|\boldsymbol{v})$ that can be formulated recursively as follows [70]

$$F(\boldsymbol{x}|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}(F(\boldsymbol{x}|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})} \tag{4.19}$$

where $C_{x,v_j|\boldsymbol{v}_{-j}}$ is a bivariate copula distribution function.

Note that a n-dimensional multivariable density can be factorized into a number of different conditional pair-copulas based on the vine copula construction proposed by Bedford and Cooke [69]. Except regular vine structure (R-vine), there are two special types of regular vines: canonical vine (C-vine) and drawable vine (D-vine). For the C-vine, each tree has a unique node that is connected to all other nodes, and the corresponding joint PDF $f(\boldsymbol{x})$ is

$$f(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c(F(x_j|x_1, ..., x_{j-1}), F(x_{j+i}|x_1, ..., x_{j-1})) \tag{4.20}$$

In contrast, each tree in D-vine is a path and the corresponding joint PDF $f(\boldsymbol{x})$ is

$$f(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c(F(x_i|x_{i+1}, ..., x_{i+j-1}), F(x_{i+j}|x_{i+1}, ..., x_{i+j-1})) \tag{4.21}$$

where the subscript indices indicate the conditional random variables to be drawn.

Copula theory and vine copulas are useful in modeling the dependence of multivariate densities in either low or high dimensional problem. A following critical question is how to select and estimate all components of a bivariate copula model or tree structure model from limited data. The next sections discuss this issue in detail.

4.2 Statistical inference of copula dependence modeling

Given a d-dimensional density, we can decompose it into products of marginal densities and bivariate copula densities and represent this decomposition with a nested set of trees that fulfill a proximity condition. However, it is often difficult to directly identify a d-dimensional density. Instead, more commonly, only data are provided. How can we estimate a pair-copula decomposition?

Statistical inference is therefore necessary for model selection and parameter estimation. In particular, small data creates large uncertainty which introduces an extra challenge in specification of the copula dependence model. Commonly, dependence modeling consists of three principal components: tree structure, copula families and copula parameters. But in this work, the uncertainty caused by lack of data in marginals is so important that it can not be simply ignored. Consequently, the marginal families and marginal distribution parameters should also be included in specification of the dependence model. As a result, the overall uncertainties U_{all} include the following five components:

$$U_{all} = \{U_s, U_{cf}, U_{cp}, U_{mf}, U_{mp}\} \tag{4.22}$$

where U_s is uncertainty in tree structures, U_{cf} and U_{cp} are referred to as the un-

certainty in copula families and parameters as well as U_{mf} and U_{mp} represent the uncertainty in marginal distribution families and parameters. To quantify these uncertainties, statistical methods are often adopted to achieve the model selection and parameter estimation for this issue.

The model uncertainty in tree structure is, in fact, a challenging issue. This is mainly because the possible decomposition of pair-copula is potentially large, especially in high dimension. In many cases, the tree structure is typically assumed to a specified model based on the informative knowledge or experience. There are also several model selection approaches in terms of specification of tree structures, including optimal C-vine structure selection [143], Bayesian approaches for D-vine [144] and maximum spanning tree for R-vine [145]. Here we do not consider tree structure model selection. Instead, our emphasis is how to efficiently quantify the uncertainties associated with copula family selection and the corresponding parameters given a vine copula structure.

4.2.1 Copula family selection and parameter estimation

Given a specified vine copula structure, classical statistical approaches, including goodness-of-fit tests [146], independence test [147] and AIC/BIC [58] are capable of handing copula families selection when datasets are large. When

both tree structure and copula families are specified, the copula parameters can be estimated in the following ways:

- Sequential estimation: parameters are sequentially estimated starting from the top tree until the last [141, 143].
- Maximum likelihood estimation: find the parameter values that maximize the likelihood function given the observations. It is efficient but has estimated standard errors numerically challenging [148].
- Bayesian estimation: Using MCMC for the posterior estimate. The prior beliefs can be incorporated and credible intervals allow assessment of uncertainty for all parameters [144, 149].

However, these classical approaches may not be the "best" way in the case of small datasets, as discussed in previous chapters. In this work, the proposed multimodel inference methodology in Chapter 2 can be easily generalized to address the model selection issue in copula dependence modeling. The corresponding copula parameters are estimated by Bayesian inference as well. A simple bivariate example is used to illustrate the basic process and performance.

Consider a bivariate random variable $v = [E_m, v_m]$ and the correlated relationship between E_m and v_m is identified by Frank copula model with model parameter $\theta = 3$. Fig. 4.3 shows the bivariate correlated data drawn from

Frank(3) copula from 10 to 5000 dataset size. Given the limited 10 data shown in Fig. 4.3 (a), it is impossible to identify a precise statistical dependence relationship. In other words, the large uncertainty resulting from the small datasets leads to a challenge in model selection and parameter estimation.

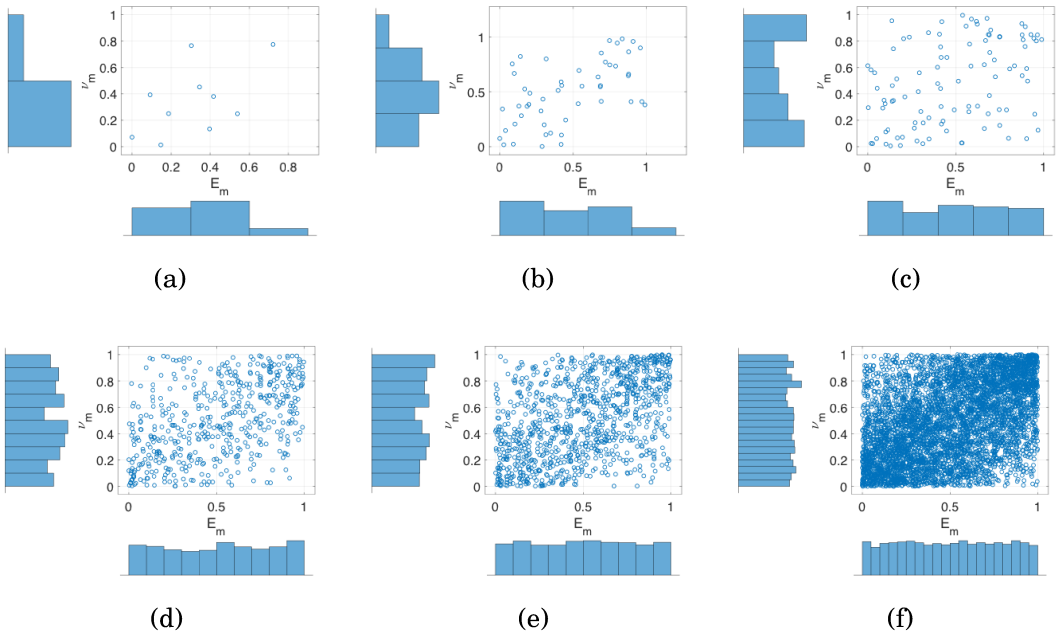


Figure 4.3: Bivariate correlated data drawn from Frank copula with copula parameter $\theta = 3$ for (a) 10 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data and (f) 5000 data

To address this issue, the multimodel inference is applied here to quantify the copula family uncertainties. Five copula models, the Gaussian, Student-t, Clayton, Gumbel and Frank copulas, are selected as the candidate copula families in this example. Bayesian multimodel inference, presented in Chapter 2, is employed here to calculate the copula model probabilities given small datasets. Without informative prior information, all candidate copula models

are assumed to have equal probability. The Monte Carlo method is adopted to compute the evidence based on Eq. (2.17), which is the integral of the copula likelihood function. Then the posterior copula model probabilities are obtained using Eq. (2.11). Fig. 4.4 shows the copula family probabilities for each candidate copula model as a function of dataset size. Notice that the model probability for the Frank copula becomes gradually larger as the dataset size increases but the multimodel inference does not select the correct Frank copula model conclusively until 1000 correlated data.

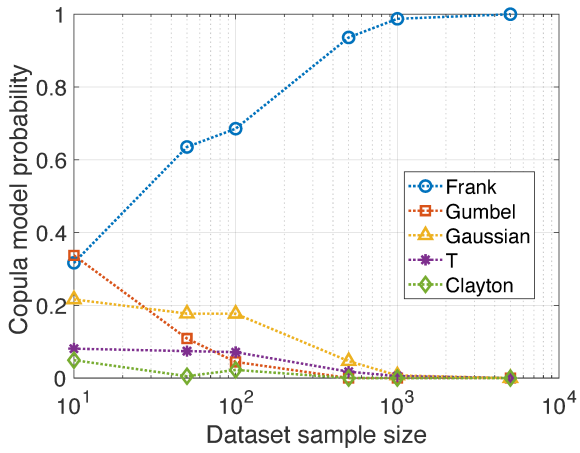


Figure 4.4: Copula model probability as a function of dataset size

Next, Bayesian inference is employed to estimate the copula parameter for the candidate models. For illustration, we again select the Frank(3) copula model as a representative example. Fig. 4.5 shows the posterior probability density for copula parameter θ for increasing dataset size. Note that a large range in parameter appears when dataset size is small and the estimate will gradually narrow with the increasing dataset size. Finally, the posterior den-

sity with 5000 data shows a very narrow distribution that is centered around the true value ($\theta = 3$).

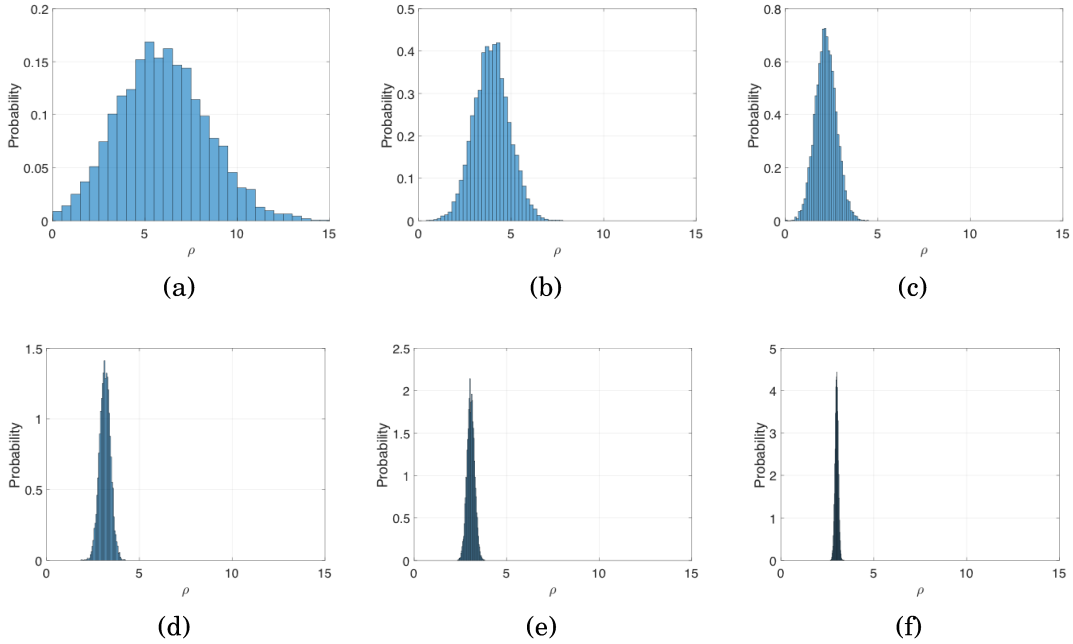


Figure 4.5: MCMC posterior copula parameter densities for the Frank copula model given (a) 10 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data and (f) 5000 data.

This simple example illustrates the multimodel inference methodology for model selection and parameter estimation of copula dependence modeling. The following section will focus on quantifying the uncertainty in marginal distributions.

4.2.2 Uncertainty in marginal distributions

In particular, as observed in the previous chapters, the marginal distribution plays a critical role in the context of small datasets. Let's consider the bivariate case as an example, where the joint pdf can be expressed as:

$$f(x_1, x_2) = c_{12}(F_1(x_1, \theta_1), F_2(x_2, \theta_2), \theta_c) \cdot f_1(x_1, \theta_1) \cdot f_2(x_2, \theta_2) \tag{4.23}$$

The joint probability density $f(x_1, x_2)$ is computed based on Eq. (4.23) with a specified copula model form c_{12}, copula model parameter θ_c, and marginal model forms f_1 and f_2 with corresponding marginal parameters θ_1 and θ_2 (note that the cdfs $F_1(x_1, \theta_1)$ and $F_2(x_2, \theta_2)$ can be simply calculated when the pdfs $f_1(x_1, \theta_1)$ and $f_2(x_2, \theta_2)$ are given). Given this expression of the joint density, it is clear that the copula model is conditional on the marginals and their parameters, which as we saw in the previous chapters have very large uncertainties in small data case. Consequently, it is necessary to identify copula model probabilities and copula parameter probabilities for each contribution of candidate marginals in the set of plausible marginal distributions. This induces a hierarchy of probabilities as illustrated in Fig. 4.6. As a result, the number of copula models grow incredibly large.

However, this growth in the number of candidate joint models only affects uncertainty propagation through the definition of the optimal sampling density

(which as we will see retains its mixture distribution construction) and in the number of importance sampling reweightings that must be performed. That is, the optimal sampling density simply contains more terms in the mixture and there are more distributions to which the samples drawn from the optimal sampling density must be reweighted. These can, in a sense, be considered as second-order effects in that they do not increase the number of model evaluations necessary for propagation.

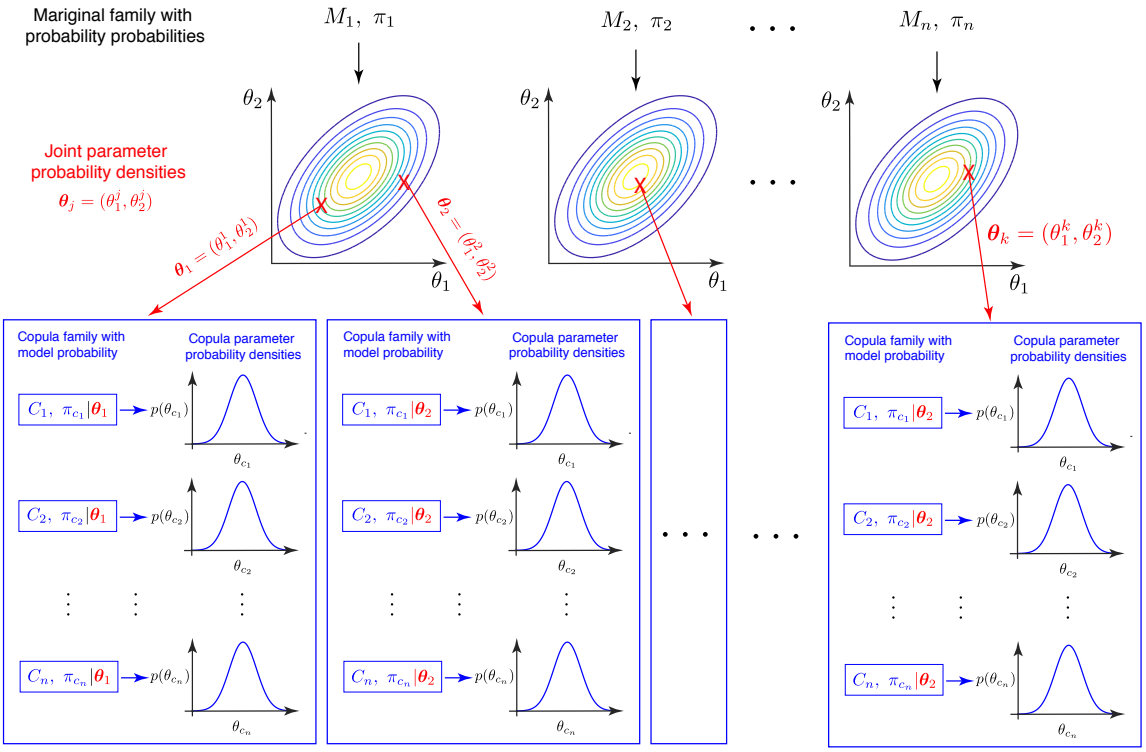


Figure 4.6: Hierarchy of Bayesian multimodel inference for copulas and marginals

4.3 Uncertainty propagation with copula dependence modeling

In Chapter 2, we proposed an efficient propagation algorithm for imprecise probabilities resulting from small datasets. It is not difficult to extend this algorithm to address the propagation of uncertainties under copula dependence modeling. Let's select the bivariate dependence modeling to illustrate this method.

4.3.1 Importance sampling for bivariate joint probability density

Consider the bivariate generic performance function $g(\boldsymbol{X}_1, \boldsymbol{X}_2)$ defining the response quantity of interest for a mathematical or physical system. The aim of uncertainty propagation is to evaluate the expectation $E(g(\boldsymbol{X}_1, \boldsymbol{X}_2))$ where $(\boldsymbol{X}_1, \boldsymbol{X}_2) \in \Omega$ is a random vector having bivariate joint probability density $p(\boldsymbol{X}_1, \boldsymbol{X}_2)$. The classical Monte Carlo estimator is computed as follows:

$$E_p[g(\boldsymbol{X}_1, \boldsymbol{X}_2)] = \int_{\Omega} g(\boldsymbol{x}_1, \boldsymbol{x}_2) p(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2 \approx \frac{1}{N} \sum_{i=1}^{N} g(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i) \tag{4.24}$$

where $E_p[\cdot]$ is the expectation with respect to $p(\cdot)$ and $(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)$ are bivariate random samples drawn from $p(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Importance sampling allows samples to be drawn from an alternative bivariate joint density $q(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and then reweight the samples to obtain the estimator. The Monte Carlo estimator in Eq. (4.24) is then modified as:

$$\begin{aligned} E_q\left[g(\boldsymbol{X}_1, \boldsymbol{X}_2)\frac{p(\boldsymbol{X}_1, \boldsymbol{X}_2)}{q(\boldsymbol{X}_1, \boldsymbol{X}_2)}\right] &= \int_\Omega g(\boldsymbol{x}_1, \boldsymbol{x}_2)\frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{q(\boldsymbol{x}_1, \boldsymbol{x}_2)}d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\ &\approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)\frac{p(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)}{q(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)} \end{aligned} \tag{4.25}$$

where $E_q[\cdot]$ is the expectation under $q(\cdot)$. The importance weights are defined as:

$$w(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i) = \frac{p(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)}{q(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)} \tag{4.26}$$

which is very important in the proposed algorithm.

4.3.2 Optimal important density for bivariate joint probability density

An optimal sampling density for propagating multiple distributions has been proposed in the previous Chapter 2. This was achieved by defining an optimization problem to minimize the mean square difference (MSD) between the set of target densities and sampling density. Eq. (2.42) and Eq. (2.43) in

Chapter 2 give a general expression for the optimal sampling density. It is straightforward to generalize this expression from the one-dimensional case to multivariate joint probability densities. If the bivariate joint density is independent, the optimal sampling density is expressed as:

$$\hat{q}(\boldsymbol{x}) = \frac{1}{N_{d_1} N_{d_2}} \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{N_{d_2}} E_\theta \left[p_{ij}(\boldsymbol{x}|\boldsymbol{\theta}) \right] \tag{4.27}$$

and the bivariate joint density $p_{ij}(\boldsymbol{x}|\boldsymbol{\theta})$ can be decomposed by marginal distribution $f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ and $f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2)$ as follows:

$$p_{ij}(\boldsymbol{x}|\boldsymbol{\theta}) = f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \cdot f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \tag{4.28}$$

where N_{d_1} and N_{d_2} are the number of candidate probability models for the marginal densities respectively and $N_d = N_{d_1} \cdot N_{d_2}$ is the total number of candidate probability models for the bivariate joint density. Thus, the optimal sampling density for independent bivariate joint density is further expressed as:

$$\begin{aligned}
\hat{q}(\boldsymbol{x}) &= \frac{1}{N_{d_1} N_{d_2}} \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{N_{d_2}} E_\theta \left[f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right] \\
&= \frac{1}{N_{d_1} N_{d_2}} \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{N_{d_2}} E_{\theta_1} \left[f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \right] E_{\theta_2} \left[f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right] \\
&= \frac{1}{N_{d_1} N_{d_2}} \sum_{i=1}^{N_{d_1}} E_{\theta_1} \left[f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \right] \sum_{j=1}^{N_{d_2}} E_{\theta_2} \left[f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right]
\end{aligned} \tag{4.29}$$

Again, it is straightforward to show that this solution generalizes as:

$$\hat{q}(\boldsymbol{x}) = \frac{1}{N_{d_1} N_{d_2}} \sum_{i=1}^{N_{d_1}} E_{\theta_1} \left[\pi_1^i f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \right] \sum_{j=1}^{N_{d_2}} E_{\theta_2} \left[\pi_2^j f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right] \tag{4.30}$$

where π_1^i associated with marginal density $f_1(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ is the model probability for model M_i satisfying $\sum_{i=1}^{N_{d_1}} \pi_1^i = 1$ and π_2^j associated with marginal density $f_2(\boldsymbol{x}_2|\boldsymbol{\theta}_2)$ is the model probability for model M_j satisfying $\sum_{j=1}^{N_{d_2}} \pi_2^j = 1$.

If the bivariate joint density is dependent with copula modeling, we may consider additional copula density $c_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}_c)$ into the joint probability density and express the bivariate joint density as:

$$p_{ij}^k(\boldsymbol{x}|\boldsymbol{\theta}) = c_{12}^k(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}_c) \cdot f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \cdot f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \tag{4.31}$$

where $k = 1, ..., N_{d_c}$ is the number of candidate copula models. Similarly, we can derive the optimal sampling density for dependent bivariate joint density with copula models as follows:

$$\begin{aligned}
\hat{q}(\boldsymbol{x}) &= \frac{1}{N_{d_1} N_{d_2} N_{d_c}} \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{N_{d_2}} \sum_{k=1}^{N_{d_c}} E_{\theta} \left[c_{12}^k(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}_c) \cdot f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \cdot f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right] \\
&= \frac{1}{N_{d_1} N_{d_2} N_{d_c}} \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{N_{d_2}} \sum_{k=1}^{N_{d_c}} E_{\theta_c} \left[c_{12}^k(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}_c) \right] E_{\theta_1} \left[f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \right] E_{\theta_2} \left[f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right] \\
&= \frac{1}{N_{d_1} N_{d_2} N_{d_c}} \sum_{k=1}^{N_{d_c}} E_{\theta_c} \left[c_{12}^k(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}_c) \right] \sum_{i=1}^{N_{d_1}} E_{\theta_1} \left[f_1^i(\boldsymbol{x}_1|\boldsymbol{\theta}_1) \right] \sum_{j=1}^{N_{d_2}} E_{\theta_2} \left[f_2^j(\boldsymbol{x}_2|\boldsymbol{\theta}_2) \right]
\end{aligned} \tag{4.32}$$

Notice that the copula probability densities are independent from the marginal probability densities, the above expression can be decomposed into three formulations

$$\hat{q}(\boldsymbol{x}) = \hat{c}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \hat{f}_1(\boldsymbol{x}_1) \cdot \hat{f}_2(\boldsymbol{x}_2) \tag{4.33}$$

where

$$\hat{c}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{N_{d_c}} \sum_{k=1}^{N_{d_c}} E_{\theta_c} \left[c_{12}^k(\boldsymbol{x}_1, \boldsymbol{x}_2 | \boldsymbol{\theta}_c) \right] \tag{4.34}$$

$$\hat{f}_1(\boldsymbol{x}_1) = \frac{1}{N_{d_1}} \sum_{i=1}^{N_{d_1}} E_{\theta_1} \left[f_1^i(\boldsymbol{x}_1 | \boldsymbol{\theta}_1) \right] \tag{4.35}$$

$$\hat{f}_2(\boldsymbol{x}_2) = \frac{1}{N_{d_2}} \sum_{j=1}^{N_{d_2}} E_{\theta_2} \left[f_2^j(\boldsymbol{x}_2 | \boldsymbol{\theta}_2) \right] \tag{4.36}$$

Using these optimal sampling densities presented in Eq. (4.34) - Eq. (4.36), we can perform the importance sampling described in the previous section

$$\begin{aligned} E_{\hat{q}} \left[g(\boldsymbol{X}_1, \boldsymbol{X}_2) \frac{p(\boldsymbol{X}_1, \boldsymbol{X}_2)}{\hat{q}(\boldsymbol{X}_1, \boldsymbol{X}_2)} \right] &= \int_\Omega g(\boldsymbol{x}_1, \boldsymbol{x}_2) \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\hat{q}(\boldsymbol{x}_1, \boldsymbol{x}_2)} d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\ &= \int_\Omega g(\boldsymbol{x}_1, \boldsymbol{x}_2) \frac{c_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) f_1(\boldsymbol{x}_1) f_2(\boldsymbol{x}_2)}{\hat{c}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) \hat{f}_1(\boldsymbol{x}_1) \hat{f}_2(\boldsymbol{x}_2)} d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\ &= \int_\Omega g(\boldsymbol{x}_1, \boldsymbol{x}_2) \hat{w}_c(\boldsymbol{x}_1, \boldsymbol{x}_2) \hat{w}_1(\boldsymbol{x}_1) \hat{w}_2(\boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\ &\approx \frac{1}{N} \sum_{i=1}^{N} g(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i) \hat{w}_c(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i) \hat{w}_1(\boldsymbol{x}_1^i) \hat{w}_2^i(\boldsymbol{x}_2^i) \end{aligned} \tag{4.37}$$

where $\hat{w}_c$ is the importance weight for copula model as well as $\hat{w}_1$ and $\hat{w}_2$ are importance weights for the marginal densities. Using the expression in Eq.

(4.37), the statistical response can be estimated through the importance sampling reweighting algorithm. The derivations and expressions derived herein for bivariate joint densities can be also generalized to n-dimensional joint densities.

4.3.3 Propagation of imprecise probabilities with copula dependence modeling

With the constituents outlined in the previous chapter, the importance sampling reweighting with bivariate copula dependence structure is summarized here and a flowchart is shown in Fig. 4.7.

- Step 1: Identify the marginal modeling - Given initial data, first identify candidate marginal families $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ and compute marginal model probabilities $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ using Bayesian multimodel inference. Then estimate the copula parameter density $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ given each copula model.

- Step 2: Construct combinations of marginals - Randomly draw m subsamples from the marginal modeling to establish m combinations of marginal distributions $\{f_1^i(\boldsymbol{x}_1|\boldsymbol{M}_1, \boldsymbol{\pi}_1, \boldsymbol{\theta}_1), f_2^i(\boldsymbol{x}_2|\boldsymbol{M}_2, \boldsymbol{\pi}_2, \boldsymbol{\theta}_2), i = 1, ..., m\}$.

- Step 3: Copula modeling - identify copula modeling given each combination of marginals. Compute copula model probability $\boldsymbol{\pi}_c$ for each candi-

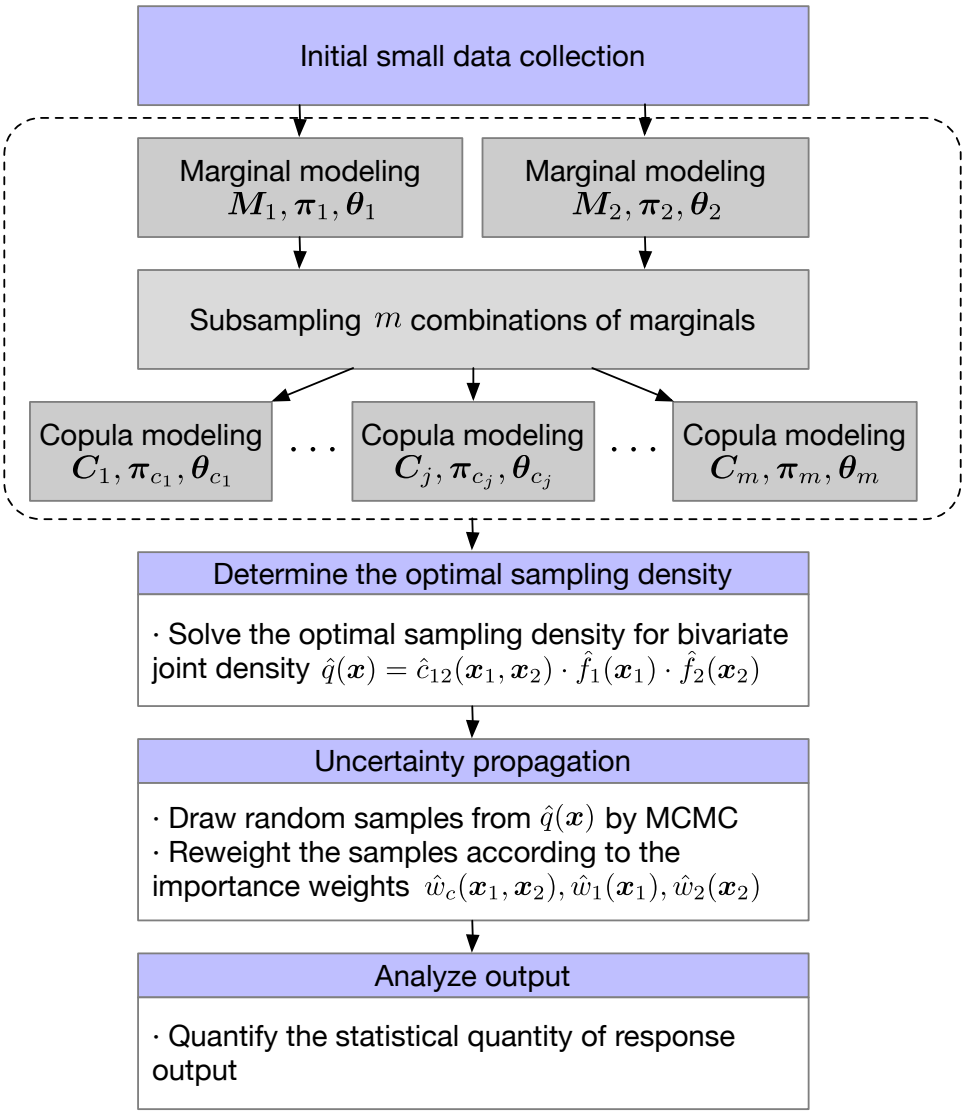


Figure 4.7: Flowchart for propagation of imprecise probabilities with copula dependence modeling

date copula model C_1 and estimate the copula parameter density θ_c given the copula model C_1.

- Step 4: Determine the optimal sampling density - Combine all the target joint densities identified from marginal modeling and copula modeling in

Step 1-3. Solve the optimal sampling density $\hat{q}(\boldsymbol{x}) = \hat{c}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \hat{f}_1(\boldsymbol{x}_1) \cdot \hat{f}_2(\boldsymbol{x}_2)$ for bivariate joint density by optimization.

- Step 5: Uncertainty propagation - Uncertainty is propagated using importance sampling with optimal sampling density $\hat{q}(\boldsymbol{x})$. Samples are drawn from $\hat{q}(\boldsymbol{x})$ using MCMC algorithm and are reweighted according to the importance weights $\hat{w}_c(\boldsymbol{x}_1, \boldsymbol{x}_2)$, $\hat{w}_1(\boldsymbol{x}_1)$ and $\hat{w}_2(\boldsymbol{x}_2)$.

- Step 8: Analyze output - Quantify the statistical response output, i.e. mean, standard deviation, etc.

4.4 Application to probabilistic prediction of unidirectional composite lamina properties

This chapter aims to apply the proposed methodology for probabilistic prediction of unidirectional composite lamina properties. We first present the basic problem description of a composite material problem and then conduct uncertainty quantification and propagation to predict the composite lamina properties.

4.4.1 Problem description

Due to their attractive properties, fiber reinforced composite material have been widely used in many industrial domains including automotive, naval, aeronautic, etc. Therefore, the evaluation of their mechanical properties is critical for the use of this type of composites. Various analytical and numerical methods are employed to evaluate the elastic properties based on the prediction of the elastic properties of unidirectional composite materials with long fibers composites [150]. These unidirectional composites are often identified as transversely isotropic materials composed of two phases: a matrix phase and a reinforcement phase, as shown in Fig. 4.8. Generally speaking, isotropic materials (e.g. epoxy), are selected as the matrix phase. Meanwhile, isotropic (e.g. glass fibers) or anisotropic (e.g. carbon fibers) fibers can be used for the reinforcement phase.

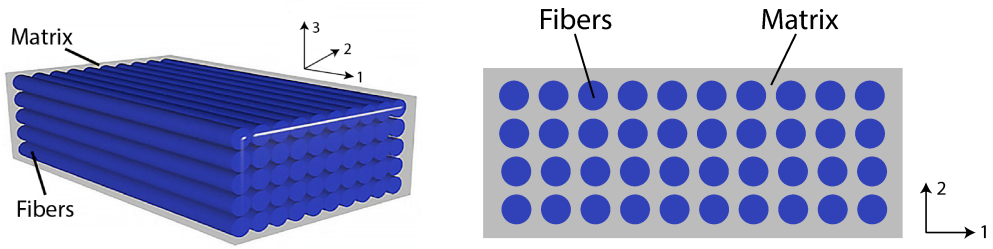


Figure 4.8: Unidirectional fiber reinforced composite (a) 3D plot and (b) 2D plot

The mechanical properties (stiffness and compliance) of transversely fiber reinforced composites are determined by six independent engineering constants

which are presented in the following compliance matrix.

$$C = \begin{bmatrix} 1/E_{11} & -\nu_{12}/E_{11} & -\nu_{12}/E_{11} & 0 & 0 & 0 \\ -\nu_{12}/E_{11} & 1/E_{22} & -\nu_{23}/E_{22} & 0 & 0 & 0 \\ -\nu_{12}/E_{11} & -\nu_{23}/E_{22} & 1/E_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/G_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/G_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/G_{12} \end{bmatrix} \tag{4.38}$$

where E_{11} and E_{22} are the longitudinal and transverse Young's moduli respectively, G_{12} and G_{23} are the longitudinal and transverse shear moduli, ν_{12} is the major Poisson's ratio and ν_{23} is the minor Poisson's ratio. Note that the stiffness matrix K is the inverse of the compliance matrix C.

The effective elastic properties are estimated based on the mechanical properties of the matrix and fibers. Several available analytical micromechanical models are useful for the prediction of the mechanical properties. Investigated models can be roughly categorized as homogenization models, elasticity models, semi-empirical models and phenomenological models. A comparative review of analytical modeling approaches can be found in [150]. The analytical approach is computational efficiently but sometimes it may not provide accurate prediction given specified conditions or assumptions. As a more general way, numerical finite element (FE) modeling is used to deal with the appropri-

ate boundary, symmetric and periodiicity conditions required to evaluate the elastic properties of unidirectional composites.

In this work, we focus on an E-Glass fiber with LY556 Polyester Resin composite material. There are five independent material properties given by Table 4.4. And the other seven dependent material properties include the fiber Young's moduli along 2 direction and 3 direction are given by

Table 4.4: Material properties of E-Glass fiber/LY556 Polyester Resin composite material model

Material property	Physical meaning	Mean value	Coefficient of variation
V_f	Fiber volume fraction	0.6	5%
E_m	Matrix's Young's modules	3.375	5%
ν_m	Matrix Poisson's ratio	0.35	5%
E_{1f}	Fiber Young's modules along 1 direction	73.01	5%
ν_{12f}	Fiber Poisson's ratio along 1-2 direction	0.228	5%

$$E_{2f} = E_{1f}, \quad E_{3f} = E_{2f} \tag{4.39}$$

and fiber Poisson's ratio along 1-3 direction and 2-3 direction are given by

$$\nu_{13f} = \nu_{12f}, \quad \nu_{23f} = \nu_{12f} \tag{4.40}$$

as well as fiber shear modulus along 1-2 direction, 1-3 direction and 2-3 direction given by

$$G_{12f} = \frac{E_{1f}}{2(1+\nu_{12f})}, \quad G_{13f} = G_{12f}, \quad G_{23f} = \frac{E_{2f}}{2(1+\nu_{23f})} \tag{4.41}$$

Using these material properties for the matrix and the fibers, the numerical FE modeling approach is adopted herein to evaluate the elastic mechanical properties of unidirectional fiber reinforced composite material. The numerical FE model is constructed as a 3D unit cell with two symmetry plan in the x-z and x-y directions and periodic boundary conditions, as shown in Fig. 4.9. This FE model has 22750 nodes and 20448 C3D8R elements in Abaqus.

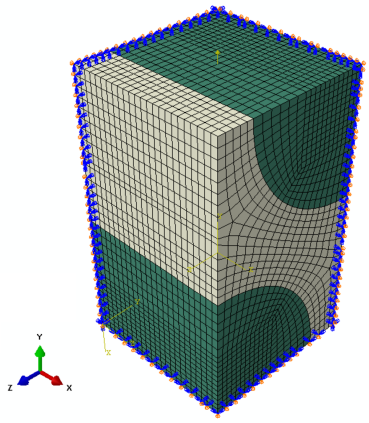


Figure 4.9: Hexagonal unit cell model

Given specified values for the material properties of matrix and fibers, the conventional study aims to explore a deterministic prediction of the elastic properties of unidirectional fiber reinforced composites. However, these specified or assumed values, ignoring their inherent variability, are often subjective or even inaccurate in many cases. Therefore, it is necessary to conduct a probabilistic prediction to well understand the influence of uncertainties associated with each matrix and fiber property on the overall property of unidirectional composites. This motivates us to develop a systematic probabilistic UQ frame-

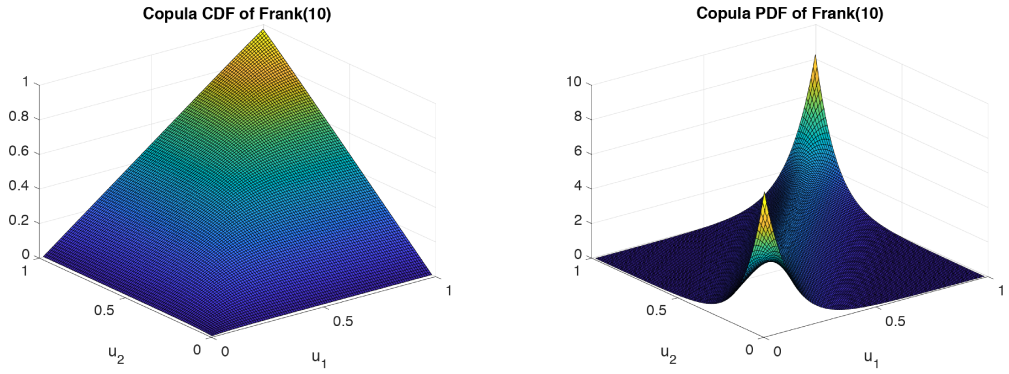
work to investigate the variabilities rooted in the composite material.

4.4.2 Probabilistic prediction of composite properties

According to the engineering experience, the five variables in Table 4.4 may be correlate or dependent and thus one task is to identify the dependence among these five variables. Commonly, the matrix properties E_m and ν_m are considered to be dependent and the fiber properties E_{1f} and ν_{12f} are dependent. But the fiber volume fraction is often assumed independent. Even though the extension to the full five variable case is straightforward, we herein focus on the influence of material property pair, E_m and ν_m for clarity and brevity in demonstration.

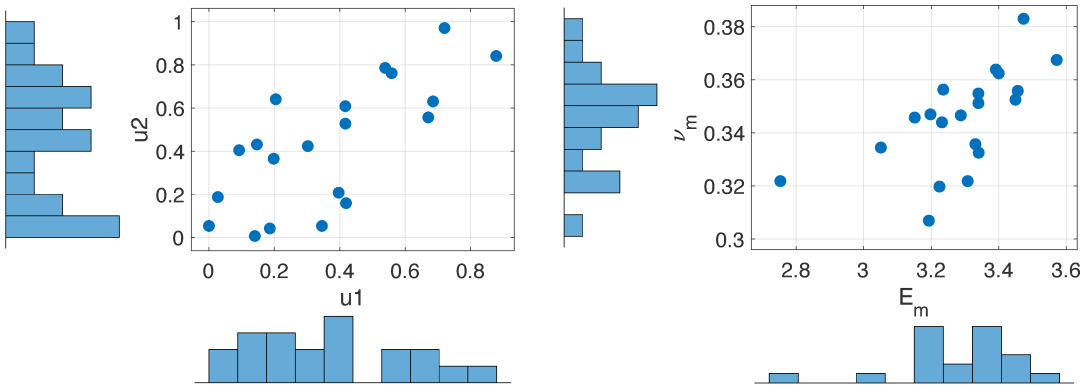
According to the nominal mean value for matrix material property is $E_m = 3.375$ and $\nu_m = 0.35$, we define a normal distribution with their nominal mean value with 5% coefficient of variation as the "true" marginal distribution for each of matrix material property. For the dependence between E_m and ν_m, we assume a Frank copula with parameter $\theta = 10$ (the corresponding Kendall's τ is 0.67), as the "true" copula to model its strong correlation. Fig. 4.10 shows the copula CDF and copula PDF for Frank(10). Using the true copula and marginals, we draw an initial 20 data from the bivariate joint density. Fig.

4.11(a) and Fig. 4.11(b) show the initial data for copula model and joint density respectively. Obviously, either copula model or marginal model cannot be precisely identified from these limited data.



(a) (b)

Figure 4.10: Frank(10) copula model (a) CDF (b) PDF

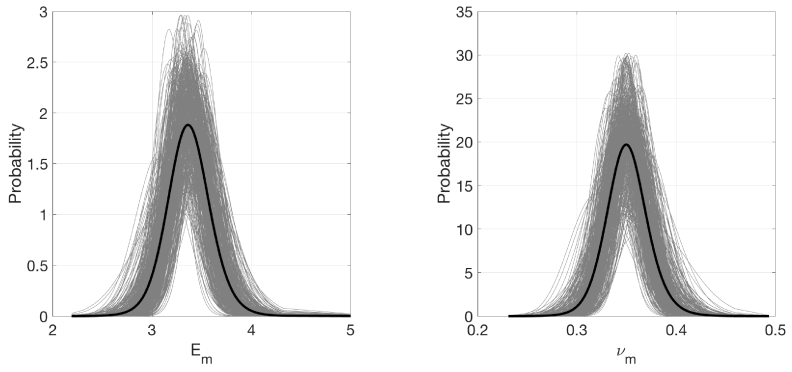


(a) (b)

Figure 4.11: 20 randomly generated matrix material properties that serve as the initial dataset (a) copula data (b) $E_m - \nu_m$ marginal data.

According to the methodology flowchart, shown in Fig. 4.7, we first consider marginal modeling to quantify the uncertainties resulting from each limited marginal data. Gaussian, Gamma, Lognormal and Weibull distributions are

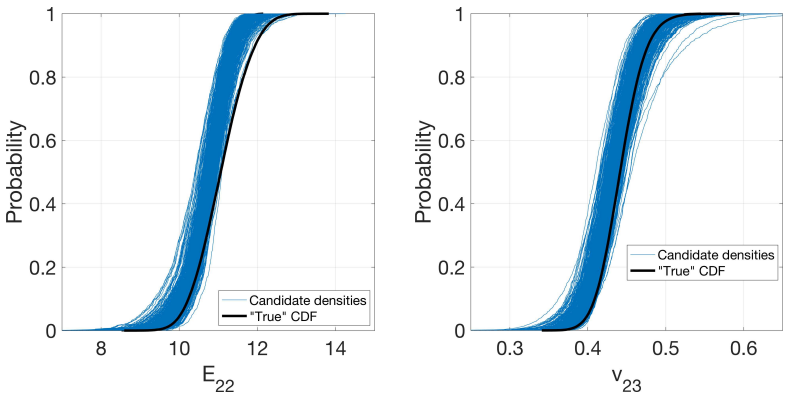
selected as the candidate probability models for marginal modeling. Bayesian multimodel inference is employed to estimate the model-form uncertainty for marginals. For each candidate model, model parameter uncertainty is estimated by Bayesian inference using MCMC algorithm. We then obtain a cloud of candidate densities for each marginal, as shown in Fig. 4.12. If the bivariate joint density is assumed independently, we randomly draw 500 subsamples as the combinations of two marginals and determine the optimal sampling density based on Eq. (4.30). The uncertainties are then propagated by importance sampling reweighting. Fig. 4.13 shows the collection of candidate empirical CDFs for each composite property E_m and ν_m given the initial 20 data. The "True" CDF in Fig. 4.13 (black curves) is presented based on the true normal marginal distributions and true copula model Frank(10).



(a) (b)

Figure 4.12: Multiple candidate probability densities for marginals (a) E_m and (b) ν_m

If the bivariate joint density is assumed dependently, copula modeling will



(a) (b)

Figure 4.13: Collection of candidate empirical CDFs for composite property (a) E_m and (b) ν_m given 20 data

be introduced as described in Fig. 4.7. For a specified combination pair of marginals, we estimate the uncertainties associated with copula modeling. There are four candidate copula models, Gaussian copula, Clayton copula, Frank copula and Gumbel copula in this example. Using the hierarchy of Bayesian multimodel inferece, we can estimate the copula model probabilities and the corresponding copula parameter densities. We then establish an ensemble of copula model sets by randomly selecting the copula models and copula parameters. Consequently, the optimal sampling density in Eq. (4.33) is determined and employed for propagation of multiple candidate densities with copula modeling. Fig. 4.14 shows the cloud of empirical CDFs with copula modeling given a specified combination of marginals. In other words, we will see the uncertainties only from copula modeling by ignoring the marginal uncertainty. Notice that the band of copula cloud is not as wide as marginals

such that it may have moderate impact on the overall uncertainties. Fig. 4.15 shows the total collection of candidate empirical CDFs for composite properties. Blue curves in Fig. 4.15 represent the candidate densities given independent assumption and grey curves represent the total candidate densities for all combinations of marginals with dependent copula-based modeling. Note that the uncertainties associated with dependence modeling show a wider band of the empirical "cloud" than the independent densities. This is the contribution from the uncertainties associated with copula modeling. In other words, combining copula modeling and marginal modeling provides a comprehensive quantification of the total uncertainties for the prediction of composite material properties. The true CDF identified herein is also included in the empirical cloud of CDFs.

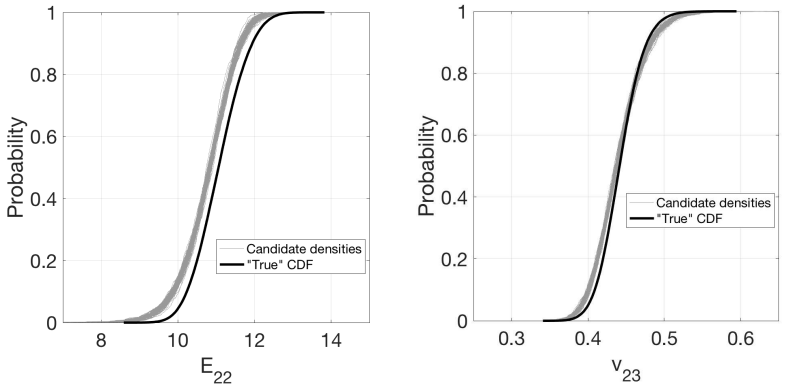
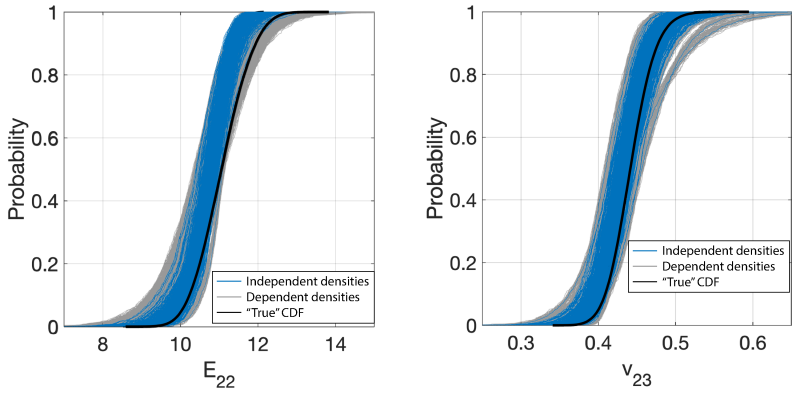


Figure 4.14: Given a specified combination of marginals, the collection of candidate empirical CDFs for composite property (a) E_m and (b) ν_m given 20 data



(a) (b)

Figure 4.15: Total collection of candidate empirical CDFs for composite property (a) E_m and (b) ν_m given 20 data

4.4.3 Influence of dataset size

In this section, we investigate the convergence of the composite material properties as a function of dataset size. As discussed in the previous chapters, small datasets led to large uncertainties including model-form (copula model and marginal model) and model parameter in the composite material properties. This raises a critical question: "How much data is necessary to gain adequate confidence in the probabilistic prediction of composite material properties?"

Here, more data are collected from 20 initial data to 50 data, 500 data and 5000 data, as shown in Fig. 4.16. The dependent trend becomes more clear as more data are collected. Table 4.5 shows the empirical CDFs for composite material properties E_{22} and ν_{23} with independent and dependent assumption as a function of dataset size from 50, 500 to 5000. Again, the empirical CDFs

with dependent modeling shows a wider band than the independent case given the same dataset size. The shape of dependent case is also slightly different from the independent case. When the dataset size reaches to 500, we note that the true CDF is not included in the cloud of independent candidate densities for E_{22}, while the true CDF is always placed in the middle of dependent candidate densities. For 5000 data, we notice a more clear difference between the true CDF and converged independent candidate densities for E_{22}, but the dependent candidate densities converges toward the true CDF. For ν_{23}, the true CDF is such close to the converged candidate densities that it is not easy to see a significant difference. The potential reason is that the dependent modeling has limited impact on the material property ν_{23} but significant influence in the material property E_{22}. As a result, it is critical to consider the dependence modeling in probabilistic prediction because simple independent assumption may lead to biased or inaccurate estimate in uncertainty quantification and propagation.

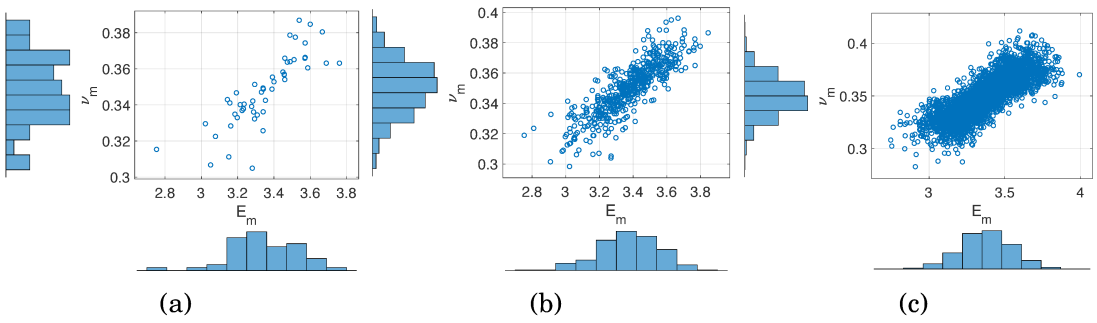
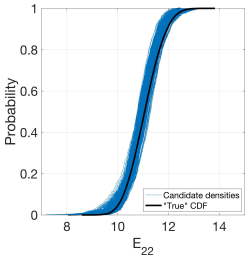
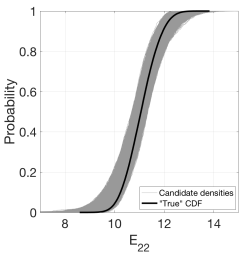
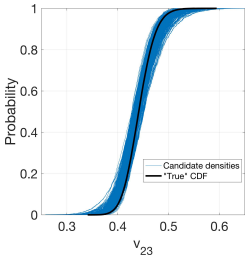
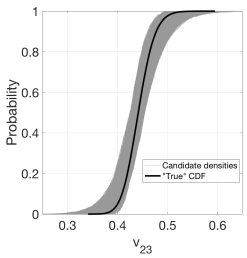
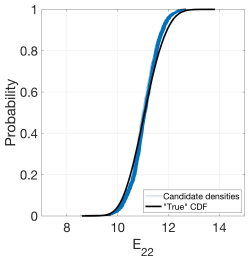
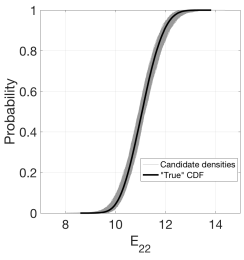
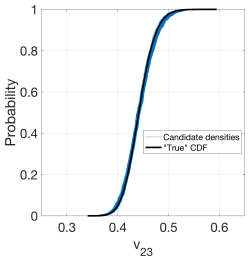
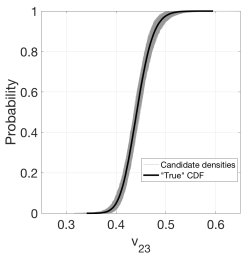
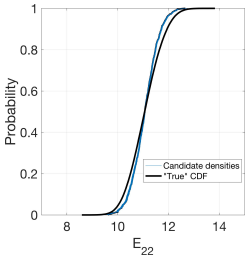
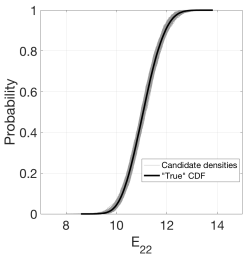
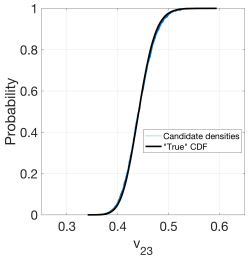
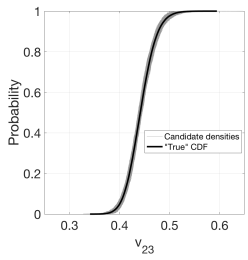


Figure 4.16: Collect dependent material property data (a) 100 data, (b) 500 data and (c) 5000 data

Table 4.5: Empirical CDFs for composite material properties with independent and dependent assumption as a function of dataset size from 50, 500 to 5000

	E_{22}		ν_{23}	
Data	Independent	Dependent	Independent	Dependent
50				
500				
5000				

4.5 Conclusion

This chapter mainly discusses the effect of dependence modeling on uncertainty quantification and propagation. Copula-based modeling methods are employed in this work to identify the dependent relationship between random variables. Then we present a systematically statistical inference including multimodel inference and parameter estimation for the copula dependence modeling. Followed by the importance sampling reweighting presented in Chap-

ter 2, we derive the optimal sampling density for bivariate joint density with independent and dependent assumption and present a description for the propagation of imprecise probabilities with copula dependence modeling. Finally, the proposed method is applied to probabilistic prediction of unidirectional composite lamina properties. The results show that the copula modeling plays an important role in accurate probabilistic prediction. Compared with simple independent modeling, the dependent copula modeling has a wider band for the empirical CDFs. With the increasing dataset size, the dependent copula modeling converges toward the true estimate, while the independent case may lead to biased or inaccurate estimate in the probabilistic prediction.

Chapter 5

Imprecise global sensitivity analysis

Global sensitivity analysis (GSA) plays an increasingly critical role in determining the critical random input parameters that drive the uncertainty of output predictions. This chapter aims at investigating how to estimate imprecise sensitivity indices given limited datasets. The multimodel methodology provided in the previous chapters is employed to quantify the uncertainties resulting from small datasets of input parameters. A set of candidate probability distributions are therefore identified as inputs for estimating sensitivity indices. Using importance sampling reweighting, these uncertain inputs are propagated through the computational model and consequently yield the imprecise sensitivity indices. An example of composite material properties pre-

diction is used to illustrate the effectiveness of the proposed methodology.

5.1 Variance-based methods for GSA

Variance-based methods are widely used for GSA. These methods decompose the variance of the output of the computational model or system into fractions which can be attributed to input variables within a probabilistic framework. Variance-based measures of sensitivity are attractive as they are applicable to the whole input space, and they can also deal with nonlinear responses and measure the effect of interactions in non-additive systems [151].

5.1.1 Sobol indices

Generally speaking, any model or system can be viewed as a function from a black-box perspective. Here we assume that $y = f(\boldsymbol{x})$ is an integrable function, where $\boldsymbol{x}$ is a vector of d uncertainty model inputs $\{x_1, ..., x_d\}$ that are mutually independent. Sobol [152] proved that $f(\boldsymbol{x})$ can be decomposed in the following way:

$$y = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \leq i < j \leq d}^{d} f_{ij}(x_i, x_j) + \cdots + f_{1,2,...,d}(x_1, x_2, ..., x_d) \tag{5.1}$$

where f_0 is a constant, $f_i(x_i)$ are univariate functions of x_i, $f_{ij}(x_j, x_j)$ are bivariate functions of (x_i, x_j) and $f_{1,2,...,d}(x_1, x_2, ..., x_d)$ are multivariate functions of $\{x_1, ..., x_d\}$. The decomposition in Eq. (5.1) is known as the high dimensional model representation (HDMR). In the HDMR, the integral of each summand $f_{1,2,...,d}(x_1, x_2, ..., x_d)$ over any of its arguments is zero:

$$\int f_{i_1,i_2,...,i_s}(x_{i_1}, x_{i_2}, ..., x_{i_s})dx_k = 0, \quad 1 \leq i_1 < ... < i_s \leq M \quad k = i_1, ..., i_s \tag{5.2}$$

This leads to definitions of the functional decomposition in terms of conditional expected values

$$f_0 = E[f(\boldsymbol{x})] \tag{5.3}$$

$$f_i(x_i) = E[f(\boldsymbol{x})|x_i] - f_0 \tag{5.4}$$

$$f_{ij}(x_i, x_j) = E[f(\boldsymbol{x})|x_i, x_j] - f_0 - f_i - f_j \tag{5.5}$$

where $E[\cdot]$ is the expectation operation. Note that f_i is the effect of varying x_i alone, also referred to as the main effect of x_i, and f_{ij} is known as the second-order interaction, that is the effect of varying x_i and x_j simultaneously.

Since the input parameters are assumed to be the mutually independent random variables, the variance of the model is defined as

$$V = \mathbf{Var}[f(\boldsymbol{x})] = \int f^2(\boldsymbol{x})d\boldsymbol{x} - f_0^2 = \sum_{s=1}^{d} \sum_{1 \leq i_1 < \cdots < i_s \leq d}^{d} \int f_{i_1,...,i_s}^2 dx_{i_1}...x_{i_s} \tag{5.6}$$

This finally leads to the following decomposition of the variance:

$$V = \sum_{i=1}^{d} V_i + \sum_{1 \leq i < j \leq d}^{d} V_{ij} + \cdots + V_{1,2\ldots,d} \tag{5.7}$$

where the partial variances are calculated as follows:

$$V_{i_1,\ldots,i_s} = \int f^2_{i_1,\ldots,i_s}(x_{i_1},\ldots,x_{i_s}) d\boldsymbol{x}, \quad 1 \leq i_1 < \ldots < i_s \leq M, \quad s = 1,\ldots,d \tag{5.8}$$

The Sobol indices are defined as the relative contribution of the partial variances to the total variance following the decomposition in Eq. (5.7)

$$S_{i_1,\ldots,i_s} = \frac{V_{i_1,\ldots,i_s}}{V} = \frac{V_{i_1,\ldots,i_s}}{\sum_{i=1}^{d} V_i + \sum_{1 \leq i < j \leq d}^{d} V_{ij} + \cdots + V_{1,2\ldots,d}} \tag{5.9}$$

such that:

$$\sum_{i=1}^{d} S_i + \sum_{1 \leq i < j \leq d} S_{ij} + \ldots + S_{1,2,\ldots,d} = 1 \tag{5.10}$$

where the index S_i measures the contribution of each variable x_i to the variance of y taken separately without interacting with any other inputs, hence S_i is common called first-order index. Higher order indices, S_{ij} in Eq. (5.10) measure the interactive contributions to the total variance. Using S_i, S_{ij} and higher order indices, one can learn the impact of each input variable in determining the output variance. However, if the model dimension is large, it needs the evaluation of $2^d - 1$ indices, which is extremely computationally intensive. A

measure of the total contribution of variable i is

$$S_i^T = \sum_{\{i\} \subset \{i_1,\ldots,i_s\}} \frac{V_{i_1,\ldots,i_s}}{V} \tag{5.11}$$

which is used to measure the contribution of variable x_i to the output variance caused by its main effects and interactions with any other input variables. It is worth noting that unlike the first order indices,

$$\sum_{i=1}^{d} S_i^T \geq 1 \tag{5.12}$$

This is because the interaction effect between, for example, x_i and x_j is contained in both S_i^T and S_j^T. The sum of the S_i^T is equal to 1 if and only if the model is purely additive without any interaction effects.

5.1.2 Estimating Sobol indices using the Monte Carlo method

The first order indices S_i in Eq. (5.10) and total indices S_i^T in Eq. (5.11) can be calculated based on the following formulations:

$$S_i = \frac{V_i}{V} = \frac{\mathbf{Var}_{x_i}[E_{x_{-i}}[y|x_i]]}{\mathbf{Var}[y]} \tag{5.13}$$

$$S_i^T = \sum_{\{i\} \subset \{i_1,\ldots,i_s\}} \frac{V_{i_1,\ldots,i_s}}{V} = \frac{E_{x_{-i}}[\mathbf{Var}_{x_i}[y|x_{-i}]]}{\mathbf{Var}[y]} \tag{5.14}$$

where Var$[\cdot]$ means the variance operation and x_{-i} means all the model inputs but not ncluding x_i. Also, S_i^T can be written as

$$S_i^T = 1 - S_{-i} = 1 - \frac{\mathbf{Var}_{x_{-i}}[E_{x_i}[y|x_{-i}]]}{\mathbf{Var}[y]} \tag{5.15}$$

where S_{-i} is referred as to the sum of all $S_{i_1,\ldots,i_s}$ other than index i.

The calculation of above indices analytically is often nontrivial unless the model or system is analytically tractable in low dimension. More commonly, sampling-based methods, i.e. Monte Carlo methods, are used to compute the Sobol indices. However, direct Monte Carlo simulation is usually unrealistic due to the large computational costs in the estimator (as shown in Eq. (5.13) and Eq. (5.14)) which requires a double-loop Monte Carlo analysis [71]. This cost is unacceptable if a single model evaluation is time-consuming because it often requires at least 1000 random samples for each Monte Carlo simulation in practice.

Many advanced algorithms, including analytical/numerical methods as well as sample-based methods, have been developed to reduce the computational cost of estimating Sobol indices. In the analytical/numerical methods, the original model is often approximated by a surrogate models which can be easily calculated analytically or numerically. Sudret [84] proposed that if the origi-

nal model can be replaced by a polynomial chaos expansion (PCE), the Sobol indices can be computed analytically from the PCE coefficients. Le Gratiet et al. [153] proposed the Gaussian process (GP) based GSA method such that the confidence intervals on sensitivity indices can be derived straightforwardly from the properties of GPs.

Compared to the analytical methods, sample-based methods are more widely applied for engineering because of their simplicity in implementation [82, 154]. Sobol [152] first proposed the following formula for calculation of V_i in the first order GSA indices (see Eq. (5.13)):

$$V_i = \int f(\boldsymbol{x}) f(x_i, \boldsymbol{\xi}_{-i}) p(\boldsymbol{x}) p(\boldsymbol{\xi}_{-i}) d\boldsymbol{x} d\boldsymbol{\xi}_{-i} - E^2[f(\boldsymbol{x})] \tag{5.16}$$

where $p(\cdot)$ is the joint probability density function (PDF) and it is the product of the PDFs of each individual random variables given the mutually independent assumption, and $\boldsymbol{\xi}_{-i}$ means all the variables in $\boldsymbol{\xi}$ not including ξ_i. Sample-based methods, i.e. the Monte Carlo method, can be used to estimate V_i in Eq. (5.16) as follows:

$$V_i \cong \frac{1}{m} \sum_{k=1}^{m} f(\boldsymbol{x}^k) f(x_i^k, \boldsymbol{\xi}_{-i}^k) - \left[\frac{1}{m} \sum_{k=1}^{m} f(\boldsymbol{x}^k) \right]^2 \tag{5.17}$$

where the subscript i is the index of the model inputs, superscript k is the index of the samples and m is the number of samples used for the estimator in Eq.

(5.17). Thus, $\boldsymbol{x}^k$ means the k-th sample of $\boldsymbol{x}$, x_i^k is the k-th sample of $\boldsymbol{x}$ for the i-th index of the model inputs, and $\boldsymbol{\xi}_{-i}^k$ is the k-th sample of $\boldsymbol{\xi}$ other than ξ_i. According to the sample-based method, Eq. (5.17) requires m samples of $\boldsymbol{x}$ and m samples of $\boldsymbol{\xi}$, which are drawn from the joint PDF $p(\boldsymbol{x})$ of the model inputs. The total model evaluations in Eq. (5.17) is therefore consisting of m evaluations of $f(\boldsymbol{x}^k)$, m evaluations of $f(x_i^k, \boldsymbol{\xi}_{-i}^k)$ and dm evaluations for all the model inputs. In other words, the overall computational cost for GSA Sobol indices is $(d+2)m$ model evaluations.

Various sample-based approaches have been proposed to improve the accuracy or reduce the computational costs of estimating Sobol indices. These approaches are usually classified into two aspects. One way focuses on the improvement of the formulation of the Sobol indices that is more accurate without introducing additional model evaluations [155–157]. Another way is to improve the efficiency of the estimator using, for example, sequences with low-discrepancy sequences, i.e. quasi-Monte Carlo methods (i.e. Sobol sequence [158]) and Latin hypercube sampling [159, 160]. In other words, given the probability distributions of the model inputs, these sampling methods using fewer samples to achieve variance reduction of the estimator.

5.2 Imprecise probability distribution given small datasets

Sample-based methods for sensitivity analysis, as presented in the previous chapter, are all initialized from the *samples* that are drawn from a specific or assumed probability distribution density. As the critical first step in GSA, probability distribution assignment is so important that it may have a strong impact on the estimation of sensitivity indices. Conventional statistical methods are well-suited for determining an appropriate probability model given large dataset size. However, large data are rarely available in engineering practice. Instead in many cases, only small datasets can be collected from experiments or simulations. For such cases, it is clearly infeasible to identify a unique probability model for the underlying probabilities of model inputs as discussed previously. Thus, the proposed multimodel methodology can be employed to address this challenge for performing GSA given lack of data.

5.2.1 Bayesian multimodel methodology

Small data creates a specific form of epistemic uncertainty which manifests in the inability to identify a single "best" probability model. Both model-form uncertainty and model parameter uncertainty are important and necessary to

consider. In this chapter, the Bayesian multimodel methodology, as shown in Chapter 2, is employed to determine the posterior model probabilities given a collection of m candidate models $\mathcal{M} = \{M_j\}$ with data $\boldsymbol{d}$.

For each of these models $M_j \in \boldsymbol{\mathcal{M}}$ there are, of course, additional uncertainties associated with model parameters $\boldsymbol{\theta}_j$. These uncertainties are quantified using Bayesian inference in Eq. (2.18). As a result, a set of candidate probability models are generated, based on the limited data case, to replace the unique probability model which is often assumed.

5.2.2 Informative prior in Bayesian framework

Given that the Bayesian multimodel methodology employed herein and the datasets used for inference are necessarily small, prior probabilities in both probability model-form and model parameters are shown to have a significant impact on quantified uncertainties [61] as discussed in Chapter 3. In Chapter 3, we employ informative priors for both model-form and model parameter uncertainties under realistic data availability constraints. The appropriate use of informative priors illustrates the power of the Bayesian approach: information gathered from previous studies, past experiences or expert opinions can be combined with new data in a natural way. But the use of an inappropriate prior, even one that seems reasonable, can lead to errors and bias that persist even as large datasets are collected.

In the context of global sensitivity analysis, particularly in engineering practice, interval bounds or variation, as one of informative source, is often provided in advance. Given probability model M_j, the corresponding interval, i.e. upper and lower bound, can be thought as the informative prior of model parameters $\boldsymbol{\theta}_j$

$$p(\boldsymbol{\theta}_j|M_j) = U(\underline{\boldsymbol{\theta}}_j, \overline{\boldsymbol{\theta}}_j) \tag{5.18}$$

where $\underline{\boldsymbol{\theta}}_j$ and $\overline{\boldsymbol{\theta}}_j$ are the lower bound and upper bound of the model parameters respectively.

The use of an appropriately informative prior serves to narrow the large uncertainty resulting from extremely sparse data.

5.3 Efficient imprecise global sensitivity analysis

The GSA methods presented above are typically defined as a measure of variability in the sense of aleatory uncertainty. In the context of epistemic uncertainty due to lack of data, imprecise probability methods can be applied. The proposed Bayesian multimodel methodology can be therefore employed to estimate the imprecise GSA Sobol indices.

As mentioned previously, the global sensitivity analysis often requires a

large sample size for computational model evaluations, and consequently, the indices estimated by the variance-based method are computationally intensive. Moreover, small data creates large uncertainties in the assignment of random input probability models. As a result, multiple probability models are identified using the Bayesian multimodel methodology to represent the imprecise probability associated with input variables. If the conventional Monte Carlo-based method is used, estimation of Sobol indices will require

$$C = N_c \cdot (d+2) \cdot m \tag{5.19}$$

model evaluations, where N_c is the total number of candidate probability models. Commonly, N_c is a large number (>5000) to fully represent the set of candidate models. In other words, the use of imprecise probability concepts increases the computational cost, probably making imprecise GSA intractable. This section aims to utilize the importance sampling reweighting, as proposed in Chapter 2, to overcome this challenge in an efficient way.

Let's revisit the formula for calculation of first-oder Sobol indices, shown in Eq. (5.16), which is rewritten in integral form as

$$V_i = \int f(\boldsymbol{x}) f(x_i, \boldsymbol{\xi}_{-i}) p(\boldsymbol{x}) p(\boldsymbol{\xi}_{-i}) d\boldsymbol{x} d\boldsymbol{\xi}_{-i} - \left(\int f(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \right)^2 \tag{5.20}$$

If importance sampling is applied here, the estimator in Eq. (5.20) can be mod-

ified by sampling from an alternative probability model as:

$$\hat{V}_i = \int f(\boldsymbol{x}) \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} f(x_i, \boldsymbol{\xi}_{-i}) \frac{p(\boldsymbol{\xi}_{-i})}{q(\boldsymbol{\xi}_{-i})} d\boldsymbol{x} d\boldsymbol{\xi}_{-i} - \left(\int f(\boldsymbol{x}) \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x} \right)^2 \tag{5.21}$$

where $q(\cdot)$ is defined as the importance sampling density and random samples $\boldsymbol{x}$ and $\boldsymbol{\xi}_{-i}$ are both drawn from $q(\cdot)$. Eq. (5.21) can be further formulated with respect to the importance reweighting given by

$$\hat{V}_i = \int f(\boldsymbol{x}) w(\boldsymbol{x}) f(x_i, \boldsymbol{\xi}_{-i}) w(\boldsymbol{\xi}_{-i}) d\boldsymbol{x} d\boldsymbol{\xi}_{-i} - \left(\int f(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x} \right)^2 \tag{5.22}$$

where $w(\boldsymbol{x}) = p(\boldsymbol{x})/q(\boldsymbol{x})$ are the importance weights, that play an essential role in the reweighting algorithm. The Monte Carlo estimator in Eq. (5.22) is then modified to correct the bias by reweighting as follows:

$$\begin{aligned} \hat{V}_i \cong & \frac{1}{m} \sum_{k=1}^{m} f(\boldsymbol{x}^k) w(\boldsymbol{x}^k) f(x_i^k, \boldsymbol{\xi}_{-i}^k) w(\boldsymbol{\xi}_{-i}^k) - \left[\frac{1}{m} \sum_{k=1}^{m} f(\boldsymbol{x}^k) w(\boldsymbol{x}^k) \right]^2 \\ \cong & E[f(\boldsymbol{x}) w(\boldsymbol{x}) f(x_i, \boldsymbol{\xi}_{-i}) w(\boldsymbol{\xi}_{-i})] - E^2[f(\boldsymbol{x}) w(\boldsymbol{x})] \end{aligned} \tag{5.23}$$

The next step is to determine how to select a proposal sampling density for estimating the imprecise GSA Sobol indices. Chapter 2 proposes to use the optimal sampling density as the proposal sampling density. This is identified by minimizing the expected mean square differences between the sampling density and the ensemble of multiple probabilities models identified by Bayesian

multimodel methodology. This process corresponds to solving an optimization problem under an isoperimetric constraint, shown in Eq. (2.37). Consequently, the optimal sampling density is derived given by

$$\hat{q}(\boldsymbol{x}) = \frac{1}{N_c} \sum_{i=1}^{N_c} \pi_i E_\theta \left[p_i(\boldsymbol{x}|\boldsymbol{\theta}) \right] \tag{5.24}$$

where N_c stands for the overall number of multiple probability models, π_i means the posterior model probability for model M_i and p_i is the target probability density with parameter $\boldsymbol{\theta}$. It is straightforward to apply this optimal sampling density to estimate Sobol indices for imprecise GSA. As the random variables are assumed to be independent, the optimal sampling density can be identified for each random variable and then multiplied to obtain the joint sampling density.

Samples are drawn from the optimal sampling density $\hat{q}(\boldsymbol{x})$ and the response of the model $f(\boldsymbol{x})$ evaluated at each sample point. The Sobol indices are reweighted according to each of the N_c sample pdfs using importance sampling as shown in Eq. (5.24).

5.4 Estimating imprecise sensitivities for composite material properties

The proposed methodology for imprecise GSA is applied to estimate the sensitivity of composite material properties to the properties of its constituent materials. Chapter 4 provides description of the E-Glass fiber/LY556 Polyester Resin composite material and the finite element model used for estimating material properties. The interest of this section aims to explore the influence of each fiber/matrix material property on the overall properties of the composite material given lack of data for constituent material properties.

5.4.1 Identification of model input distributions

To determine the sensitivity indices of each model input, it is common to assume the probability distribution based on known experience or widely used options. However, in many cases, the information or options is limited and subjective such that direct assumption is sometimes inaccurate or unreasonable. Instead, this work employs the Bayesian multimodel methodology presented in Chapter 2. Due to limited collected data, it is difficult to assign a precise distribution with accurate mean and standard deviation for each material property. We therefore provide an upper and lower bound for the mean and coefficient of

variation (COV) of each material property according to engineering experience. Table 5.1 shows the basic information for the E-Glass fiber/LY556 Polyester Resin composite material model. For each material property, a set of data are collected from various literature and technical reports [161–183]. The data are listed as follows:

- ν_m: [0.35, 0.35, 0.35, 0.35, 0.35, 0.35]
- V_f: [0.6]
- E_m: [3.5, 3.35, 3.4, 3.2, 3.45, 3.35]
- ν_{12f}: [0.22, 0.2, 0.2, 0.25, 0.25, 0.3, 0.22, 0.22, 0.22, 0.23, 0.22, 0.2]
- E_{1f}: [75.79, 74, 76, 72, 72, 72.4, 70, 72.3, 74, 72.4, 72, 72.2, 72, 72.4, 71.7, 73.1, 72.35, 76, 73, 71.5, 72, 73.1, 74, 76]

Notice that the data size of ν_m, V_f and E_m is extremely sparse ($<$10). In other words, the uncertainty with these variables is extremely large, and consequently, it is difficult to identify a narrow posterior estimate in the Bayesian setting. For such cases, the prior information in Bayesian framework will play an important role in defining the uncertainties and variabilities in the estimation, as discussed in Chapter 3. In this example, the mean and coefficient of variation (COV), as critical prior statistical information (last two columns in Table 5.1), are provided based on the engineering experience in composite material studies.

Table 5.1: E-Glass fiber/LY556 Polyester Resin composite material model

Material Property	Physical meaning	Number of data	Mean bound	COV bound
ν_m	Matrix Poissons ratio	6	[0.3, 0.4]	[5%, 10%]
V_f	Fiber volume fraction	1	[0.55, 0.65]	[5%, 10%]
E_m (GPa)	Matrixs Young's modules	6	[3, 4]	[5%,10%]
ν_{12f}	Fiber Poissons ratio long 1-2 direction	12	[0.15, 0.35]	[5%, 10%]
E_{1f}(GPa)	Fiber Young's modules along 1 direction	24	[70, 80]	[5%, 10%]

Given the data and prior information, the Bayesian multimodel inference is implemented to quantify the uncertainties associated with each material property. The following candidate probability models are considered:

- Normal distribution, Lognormal distribution, Gamma distribution, Inverse Gaussian distribution, Logistic distribution

Considering equal prior model probabilities, the first step is to determine the posterior model probabilities using Bayesian multimodel inference. Table 5.2 presents the model probabilities from the given data for each material property. Clearly, the limited data has almost no effect on the model probabilities which remain largely unchanged. We then estimate the posterior joint density of model parameters using Bayesian inference. Finally, a finite set of probability models is established by randomly selecting the model family with model probability and the associated parameter values from the posterior joint densities. According to the multiple target probability models, the optimal sampling density is therefore determined and selected for importance sampling reweighting.

Table 5.2: Model probabilities from the given data for each material property

Distribution	ν_m	V_f	E_m	ν_{12f}	E_{1f}
Normal	0.200	0.200	0.203	0.104	0.180
Lognormal	0.200	0.200	0.193	0.249	0.214
Gamma	0.200	0.200	0.205	0.194	0.205
Inverse Gaussian	0.200	0.200	0.202	0.253	0.217
Logistic	0.200	0.200	0.197	0.199	0.185

Fig. 5.1 shows the ensemble of probability models for each material property. The gray curves represent the cloud of multiple distributions and the thick black curve is the optimal sampling density. It can be observed that the cloud band is related to the dataset size. The variables with extremely sparse data show a wider band, particularly the case of fiber volume fraction V_f. But the band in ν_{12f} and E_{1f}, which have more data, are narrower. This is in accordance with the previous trend discussed in Chapter 2, the cloud of candidate densities narrows as data are added.

5.4.2 Estimating of imprecise Sobol indices

Once the identification of model input distributions is completed, the following step is to calculate the sensitivity indices through the computational model. In this example, there are 500 candidate densities for each material property and total combination of these sets of densities is equal to $500^5 = 3.125^{13}$, which is obviously prohibitive. Even though the proposed importance sampling reweighting algorithm can significantly save the computational cost, this

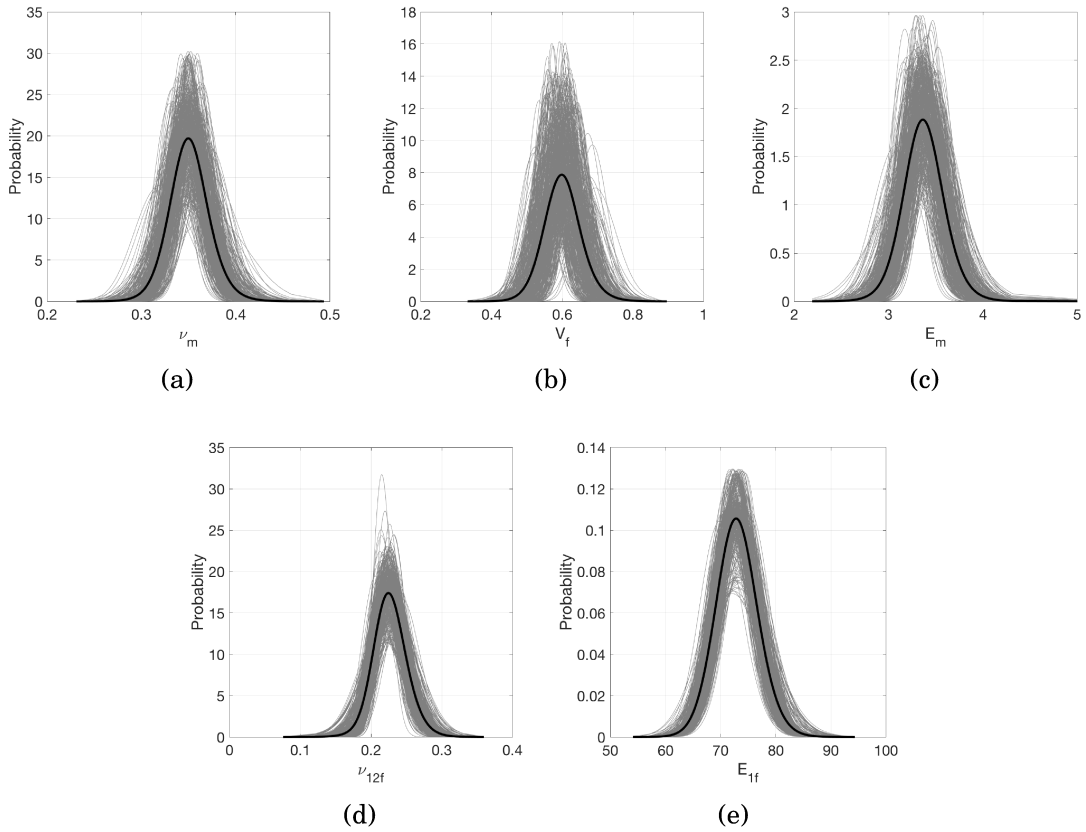


Figure 5.1: Multiple probability distributions using multimodel Bayesian methodology for (a) ν_m, (b) V_f (c) E_m, (d) ν_{12f} and (e) E_{1f}

huge number is still unacceptable. Thus, a representative 10,000 samples using Latin hypercube sampling is employed instead of the total combination. In addition, 50,000 random samples are drawn from the optimal sampling density of each material property for computational model evaluations using the finite element method as presented in Chapter 4.

There are two output composite properties, Young's modulus along 2 direction, E_2 and the Poisson's ratio in the 2-3 direction ν_{23}. Fig. 5.2 presents histograms of first-order Sobol indices for E_2. Instead of a deterministic value,

the imprecise GSA results show a probabilistic description of the sensitivity indices. This variation in sensitivity indices is caused by the uncertainties associated with the input probability models. Note that the overall influence of V_f is most significant with a range from 0.4 to 0.95. The second most significant variable is E_m which shows a variation from 0 to 0.6. ν_m with a variation range between 0 to 0.1 shows a moderate impact on the composite property E_2. The other two variables ν_{12f} and E_{1f}, meanwhile, play a very limited role in E_2. The corresponding empirical CDF results are shown in Fig. 5.3. All of these probability results clearly display the influence of uncertainties resulting from lack of data on the estimate of global sensitivity indices.

Next, let us turn the attention to another output composite property ν_{23}. From the histograms in Figure 5.4, we note that ν_m becomes the most significant variable as its Sobol indices range is from 0.8 to 1. Unlike the above case, V_f here shows a limited effect on the sensitivity of composite property ν_{23}. The other three variables, E_m, ν_{12f} and E_{1f} have such a minor impact that their influence can be effectively ignored in this case. Similarly, the empirical CDF results can be found in Fig. 5.5. Again, the histogram and CDF results systematically quantify the uncertainties and variations associated with the sensitivity indices in a probabilistic framework. In terms of the two response outputs, Table 5.3 shows the mean and standard deviation of the first-order Sobol indices for each model input parameter. The expected sensitivity of each

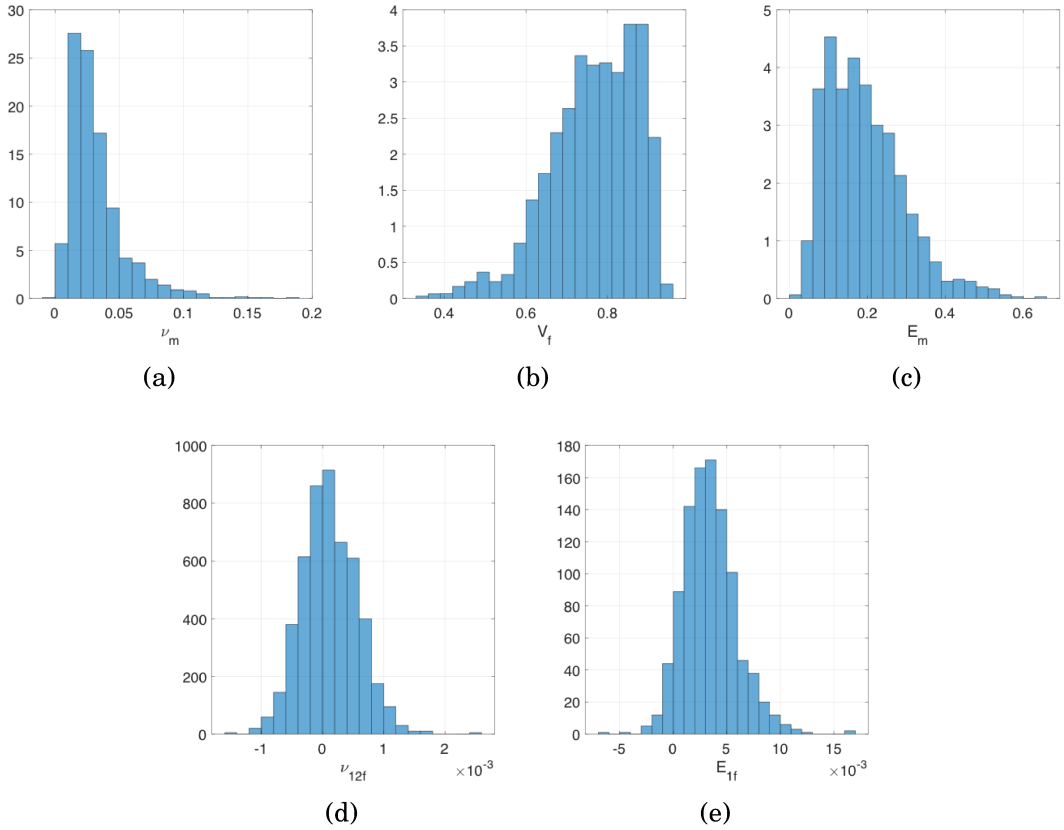


Figure 5.2: Histogram of first-order Sobol indices in terms of E_2: (a) ν_m, (b) V_f (c) E_m, (d) ν_{12f} and (e) E_{1f}

input can be evaluated by the mean and the uncertainty can be observed by the standard deviation estimator.

Table 5.3: Statistical information of GSA for two output composite properties

Output	E_2		ν_{23f}	
Parameter	Mean	Standard deviation	Mean	Standard deviation
ν_m	0.0316	0.0224	0.9421	0.0332
V_f	0.7643	0.1058	0.042	0.0300
E_m	0.1919	0.0990	0.0049	0.0035
ν_{12f}	0.0002	0.0015	0.0023	0.0025
E_{1f}	0.0034	0.0025	0.0034	0.0023

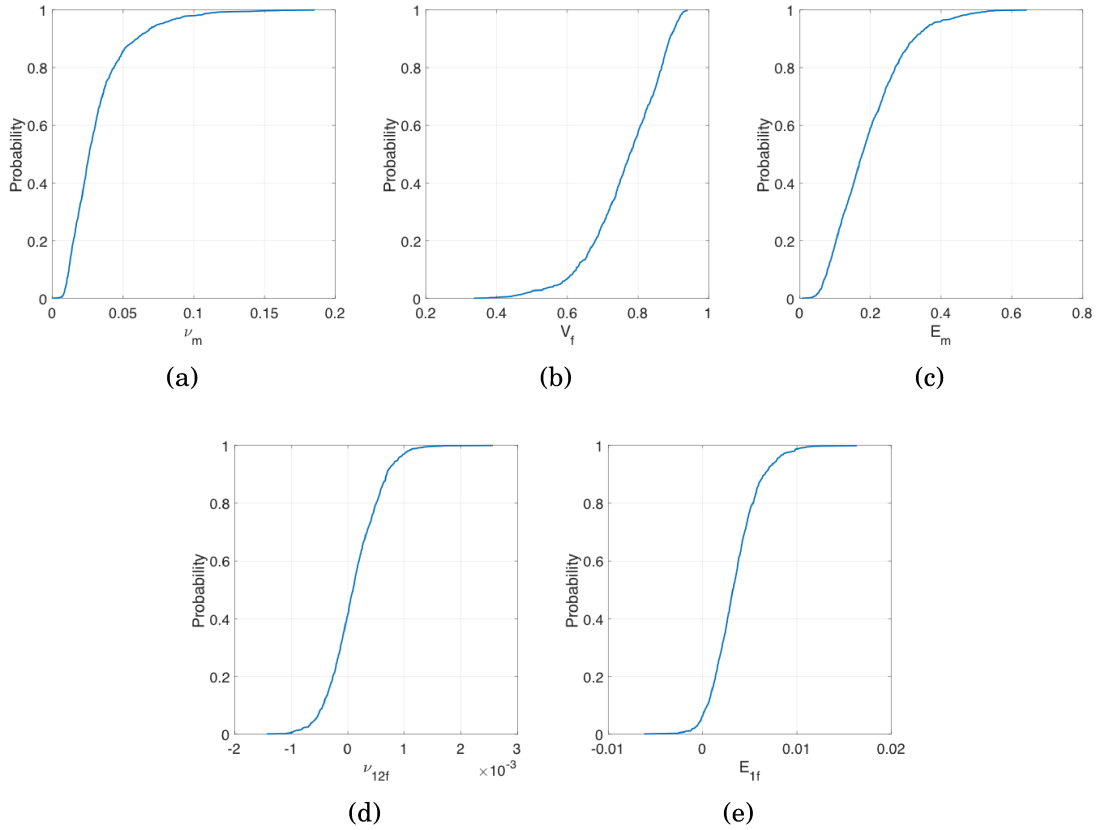


Figure 5.3: CDF of first-order Sobol indices in terms of E_2: (a) ν_m, (b) V_f (c) E_m, (d) ν_{12f} and (e) E_{1f}

5.5 Conclusion

Global sensitivity analysis aims to determine which uncertain input variables of a computational model primarily influence the output response most. Sobol indices are widely used in this context when the model inputs are random variables. Practically, input random variables are affected by both aleatory and epistemic uncertainty. The latter is often caused by lack of data. Therefore, imprecise probability representations become popular to address this issue. In

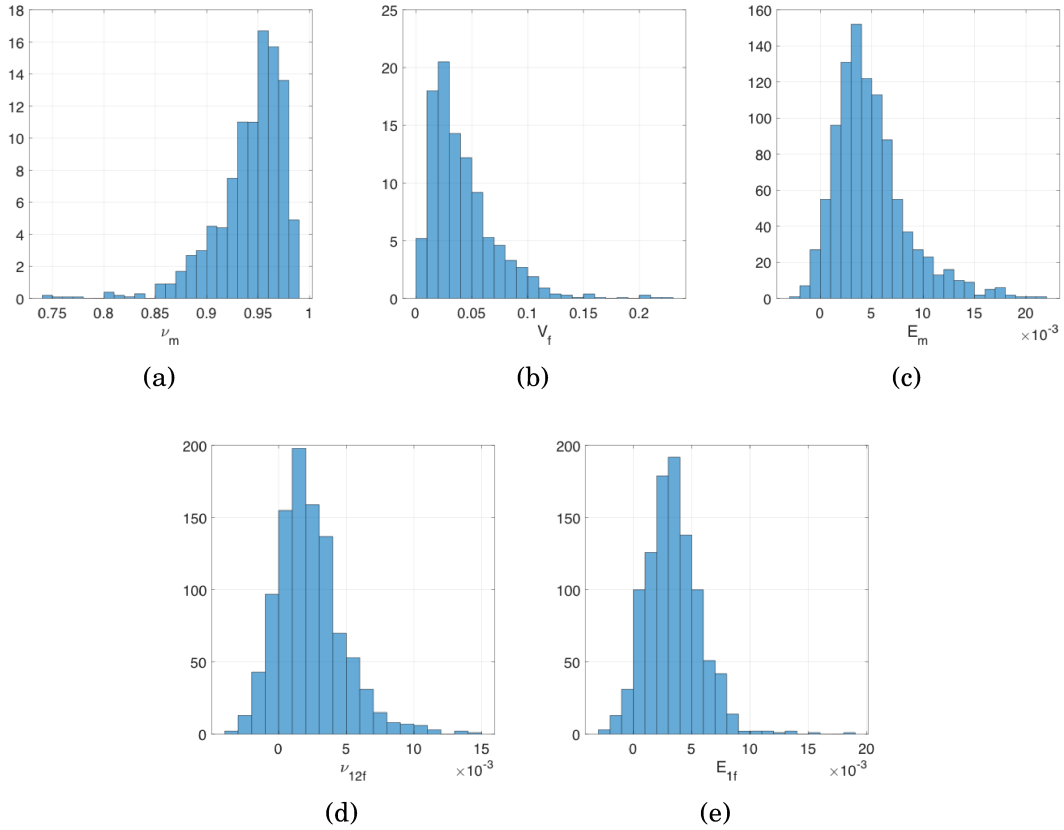


Figure 5.4: Histogram of first-order Sobol indices in terms of ν_{23}: (a) ν_m, (b) V_f (c) E_m, (d) ν_{12f} and (e) E_{1f}

this chapter, we extend the proposed Bayesian multimodel methodology and optimal importance sampling reweighting algorithm to provide an efficient estimate of imprecise Sobol indices. Instead of direct assumption of input probability models, the multimodel inference methodology is applied to quantify the model-form and model parameter uncertainties caused by small datasets and generate a set of candidate probability models for the model inputs. An original importance reweighting algorithm is proposed for calculation of imprecise first-order Sobol indices. As a result, the uncertainty associated with

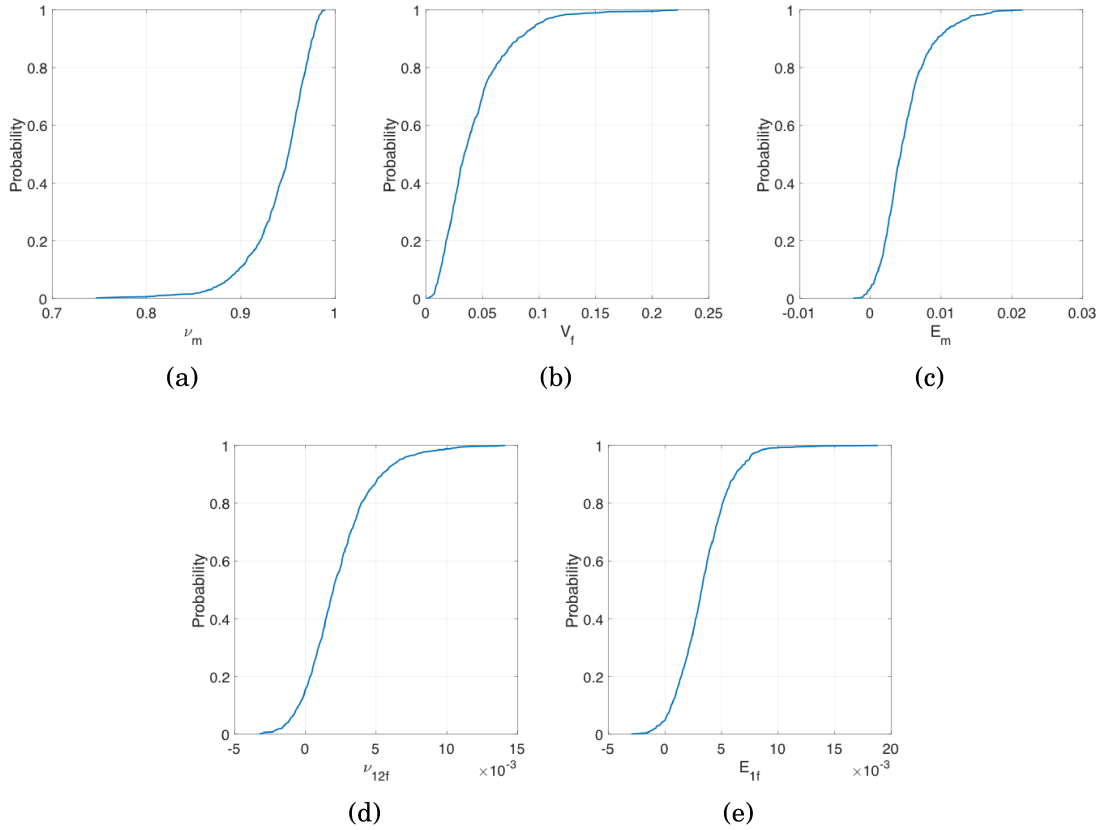


Figure 5.5: CDF of first-order Sobol indices in terms of ν_{23f}: (a) ν_m, (b) V_f (c) E_m, (d) ν_{12f} and (e) E_{1f}

sensitivity indices are systematically quantified at much lower computational cost. The approach is illustrated by a composite material problem and a fully probabilistic description of sensitivity indices is presented for the prediction of composite material properties.

Chapter 6

Conclusion and future works

This thesis presents a novel methodology for performing the uncertainty quantification and propagation from small data. There have been many studies in the past few years to quantify and propagate the epistemic uncertainties resulting from lack of data, knowledge or information from the imprecise probabilities framework. Unlike the conventional studies, this work introduces a systematical methodology to quantify the uncertainties in a probabilistic way and put the emphasis on efficiently propagating these uncertainties. The developed algorithm can also be updated to adaptively accommodate added data. This allows us to initially explore the prediction trend in small data and dig into the characteristics of convergence with gradually increased dataset size. In fact, this is not only an effective and efficient algorithm but also a valuable data-driven methodology for uncertainty quantification and propagation.

Chapter 2 provides the principal framework to address the issue of uncertainty quantification and propagation when input data for characterizing probabilistic inference are scarce. Lack of data amplifies the uncertainties in model selection and parameter estimation. Classical Monte Carlo based approaches involve multiple loops that come at very large computational costs. The proposed methodology collapses these multiples loops to a single loop through importance sampling reweighting algorithm and achieve simultaneously propagating of uncertainties associated with ensemble probability model sets each having random variable parameters. Without requiring any additional computational costs, the proposed method presents an adaptive updating as additional data are collected.

Chapter 3 presents an investigation into the effect of prior knowledge on the uncertainties resulting from small datasets. When data is limited, the prior information will play an increasingly important role in Bayesian statistical learning. As a result, instead of the commonly used noninformative prior, an informative prior probability is formulated using a data-driven method. A plate buckling strength problem is employed to illustrate that prior probabilities have a significant impact on multimodel UQ for small datasets and inappropriate priors may lead to biased and even incorrect estimate even though a large number of data are given.

The UQ studies in Chapter 2 and 3, involving the one-dimensional problem,

are conducted based on the independence assumption of random variables. The task of Chapter 4 aims at overcoming the challenge of uncertainty quantification and propagation in multivariate random variables with a dependent relationship. The proposed methodology is rigorously extended to quantify the uncertainties in both marginal modeling and copula modeling. Through an example of composite material property prediction, we notice that the copula modeling plays an important role in accurate probabilistic prediction. Compared with simple independent modeling, the dependent copula modeling has a wider band for the empirical CDFs. With the increasing dataset size, the dependent copula modeling converges toward the true estimate, while the independent case may lead to biased or inaccurate estimate in the probabilistic prediction of composite material properties.

Finally, an application of the proposed methodology in sensitivity analysis is illustrated in Chapter 5. This work differs from the classical work that identifies distributions for input variables using subjective assumption or experience. Instead, a robust multimodel inference is employed to probabilistically represent the input variables in the global sensitivity analysis (GSA). Unlike the classical GSA methods, the proposed method concludes with imprecise sensitivity indices when available data for input variables are scarce. In other words, when data is very limited, it is impossible to identify deterministic sensitivity indices, and conversely, it is more reasonable to have a probabilistic

description of sensitivity indices instead.

Future works will be developed in several aspects. One is to consider the nonparametric way instead of the parametric probability method here. This is also a limitation that has been mentioned in Chapter 2. Nonparametric method, as a more flexible approach, can address the issue of probability model selection and additional model updating. Another one is to overcome the "curse of dimensionality" in the context of UQ, which is always the challenging issue in UQ community. In particular, this issue involves the difficulty in high dimensional dependence modeling. Vine copula mentioned in Chapter 4 can be used to model the dependence structure but it is still worth to explore the issue of uncertainty quantification and propagation with vine copula modeling. Additionally, it is interesting to improve the proposed method in GSA if the input variables are not independent.

The current methodology is developed based on the Monte Carlo-based method. Beyond Monte Carlo propagation, the following work may explore the improvements using variance reduction techniques, surrogate modeling, polynomial chaos expansion and stochastic collection approaches. It is also straightforward to expand the proposed methodology for reliability analysis. The future work will combine First Order Reliability Method(FORM)/Second Order Reliability Method (SORM), multiple importance sampling (MIS) and subset simulation to investigate the probability of failure in rare event. Additionally, it is also

interesting to extend the proposed methodology for addressing more general issues involving the context of “data”, for example, missing data, dirty data and unbalanced data, etc.

Appendix A

Affine-invariant ensemble MCMC algorithm

Many of the most significant gains in the probabilistic analysis including reliability analysis have come from numerical algorithms for approximate inference, particularly MCMC, which are designed to sample from posterior probability distribution efficiently even in parameter spaces with large numbers of dimensions. The simplest and most commonly used MCMC algorithms are Metropolis-Hastings (MH) [107, 184] and Gibbs sampling [106]. In this work, we use an MH-based MCMC algorithm - Affine-invariant ensemble sampler proposed by Goodman and Weare [108].

The performance of the affine-invariant ensemble algorithm is invariant under linear transformations of the parameter space. For instance, an affine

transformation is an invertible mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ of the form $z = \alpha x + \beta$. If X has the probability density $\pi(x)$, then $Z = \alpha X + \beta$ has the density

$$\pi_{\alpha,\beta}(z) = \pi_{\alpha,\beta}(\alpha x + \beta) \propto \pi(x) \tag{A.1}$$

Ensemble MCMC has the benefit that the sampler works just as well on a high degenerate Gaussian distribution as an uncorrelated and isotropic Gaussian distribution. The basic principle can be explained that many walkers, move through parameter space; at each iteration, each walker undergoes a trial move with the step being accepted with probability. In fact, the trial move is dependent on the positions of each of the other walkers, called complementary ensemble, because these provide information about the underlying distribution. We describe the algorithm procedure herein for a target posterior distribution $p(x)$ (see also Foreman-Mackey et al. [109] for more information)

Notice that affine-invariant ensemble MCMC utilizes the position of the walkers at each step for next moving. A curving distribution, such as Gamma distribution ("banana shape"), would result in a low efficiency for MH algorithm since the trial distribution cannot be tuned throughout the parameter space. Nevertheless, due to the use of multiple walkers, the positions of walkers ensures trial steps throughout the parameter space, and thus the acceptance probability is sufficiently high associated with a much shorter autocorrelation

1: Initialize the positions of the n_c walkers, and suppose the positions of all the walkers are described by $\boldsymbol{x}(t)$ at iteration t
2: For each of the walkers $x_j(t), j = 1, 2, ..., n_c$ successively
3: Draw a random walker x_k from the complementary ensemble $\boldsymbol{x}_{[j]}(t)$.
4: Generate a random variable z from

$$g(z) \propto \frac{1}{\sqrt{Z}}, z \in \left[\frac{1}{a}, a\right] \tag{A.2}$$

5: Propose a trial step y that is called stretch move

$$y = x_k + z[x_j(t) - x_k] \tag{A.3}$$

6: Define an accepted probability

$$\alpha = \min\left(1, z^{n-1}\frac{p(y)}{p(x_j(t))}\right) \tag{A.4}$$

where n is the dimension of parameter space
7: Draw a random variable $r \sim U(0, 1)$
8: Determine the next move

$$x_j(t+1) = \begin{cases} y & \text{if } r \leq \alpha \\ x_j(t) & \text{otherwise} \end{cases} \tag{A.5}$$

9: Iterate over t from step 2 to obtain $\boldsymbol{x}(t)$

time than standard MH algorithm. Also, it is straightforward to extend the single stretch move to parallel stretch move by simultaneously advancing each walker based on the stage of the ensemble instead of evolving the walkers in series. One can therefore take advantage of generic parallelization to further improve the efficiency of this algorithm. In this paper, we use $n = 50$ walkers and set the step-size parameter $a = 2$ but in fact, a can be adjusted if the acceptance fraction is too low or too high (see [108] and [109] for further discussion).

A.0.1 Advantages over traditional MCMC algorithms

Ensemble MCMC sampler has several advantages over traditional MCMC sampling algorithms and it has excellent performance as measured by the likelihood function calls per independent sample. The major advantage of the algorithm is that it leverages an ensemble of Markov chains to adopt the proposal density through an invariant affine transformation. This greatly improves efficiency for anisotropic and degenerate densities increasing the acceptance rate at the same time maintaining sample quality, and significantly reduces the "burning" period (correlation length) of the Markov chain yielding independent samples more quickly. Another benefit is that this algorithm is great "self-tuning" such that it only requires 1 or 2 tuning parameters rather than $\sim n^2$ for most traditional MH-based MCMC algorithms in an n-dimensional parameter space. Both advantages have strong effect of greatly improving the efficiency for subset simulation in reliability analysis.

Bibliography

[1] P. E. Hess, D. Bruchman, I. A. Assakkaf, and B. M. Ayyub, "Uncertainties in material and geometric strength and load variables," *Naval engineers journal*, vol. 114, no. 2, pp. 139–166, 2002.

[2] C. G. Soares, "Uncertainty modelling in plate buckling," *Structural Safety*, vol. 5, no. 1, pp. 17–34, 1988.

[3] J. Zhang and M. D. Shields, "On the quantification and efficient propagation of imprecise probabilities resulting from small datasets," *Mechanical Systems and Signal Processing*, vol. 98, pp. 465–483, 2018.

[4] U. Rüde, K. Willcox, L. C. McInnes, H. De Sterck, G. Biros, H. Bungartz, J. Corones, E. Cramer, J. Crowley, O. Ghattas *et al.*, "Research and education in computational science and engineering," *arXiv preprint arXiv:1610.02608*, 2016.

[5] R. Ghanem, D. Higdon, and H. Owhadi, *Handbook of uncertainty quantification*. Springer, 2017.

[6] R. C. Smith, *Uncertainty quantification: theory, implementation, and applications*. Siam, 2013, vol. 12.

[7] B. Peherstorfer, K. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *arXiv preprint arXiv:1806.10761*, 2018.

[8] T. Cui, K. J. Law, and Y. M. Marzouk, "Dimension-independent likelihood-informed mcmc," *Journal of Computational Physics*, vol. 304, pp. 109–137, 2016.

[9] H. N. Najm, "Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics," *Annual review of fluid mechanics*, vol. 41, pp. 35–52, 2009.

[10] O. Le Maître and O. M. Knio, *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.

[11] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *Journal of Computational Physics*, vol. 366, pp. 415–447, 2018.

[12] D. Xiu and G. E. Karniadakis, "Modeling uncertainty in flow simulations

via generalized polynomial chaos," *Journal of computational physics*, vol. 187, no. 1, pp. 137–167, 2003.

[13] C. Soize and R. Ghanem, "Physical systems with random uncertainties: chaos representations with arbitrary probability measure," *SIAM Journal on Scientific Computing*, vol. 26, no. 2, pp. 395–410, 2004.

[14] C. Soize and C. Farhat, "A nonparametric probabilistic approach for quantifying uncertainties in low-dimensional and high-dimensional nonlinear models," *International Journal for Numerical Methods in Engineering*, vol. 109, no. 6, pp. 837–888, 2017.

[15] R. Bostanabad, B. Liang, J. Gao, W. K. Liu, J. Cao, D. Zeng, X. Su, H. Xu, Y. Li, and W. Chen, "Uncertainty quantification in multiscale simulation of woven fiber composites," *Computer Methods in Applied Mechanics and Engineering*, vol. 338, pp. 506–532, 2018.

[16] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, and M. Anitescu, "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 431–441, 2011.

[17] J. P. Van Der Sluijs, M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey, "Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the nusap system,"

Risk Analysis: An International Journal, vol. 25, no. 2, pp. 481–492, 2005.

[18] A. Der Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?" *Structural Safety*, vol. 31, no. 2, pp. 105–112, mar 2009.

[19] S. Ferson and L. R. Ginzburg, "Different methods are needed to propagate ignorance and variability," *Reliability Engineering & System Safety*, vol. 54, no. 2-3, pp. 133–144, nov 1996.

[20] P. Walley, *Statistical reasoning with imprecise probabilities*. Peter Walley, 1991, vol. 42.

[21] P. Walley, "Towards a unified theory of imprecise probability," *International Journal of Approximate Reasoning*, vol. 24, no. 2-3, pp. 125–148, 2000.

[22] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing probability boxes and Dempster-Shafer structures*. Sandia National Laboratories Albuquerque, 2002, vol. 835.

[23] S. Ferson and J. G. Hajagos, "Arithmetic with uncertain numbers: rigorous and (often) best possible answers," *Reliability Engineering & System Safety*, vol. 85, no. 1, pp. 135–152, 2004.

[24] R. Schöbi and B. Sudret, "Uncertainty propagation of p-boxes using

sparse polynomial chaos expansions," *Journal of Computational Physics*, vol. 339, pp. 307–327, 2017.

[25] A. E. Raftery, "Bayesian model selection in social research," *Sociological methodology*, pp. 111–163, 1995.

[26] S. Sankararaman and S. Mahadevan, "Distribution type uncertainty due to sparse and imprecise data," *Mechanical Systems and Signal Processing*, vol. 37, no. 1, pp. 182–198, 2013.

[27] I. Molchanov, *Theory of Random Sets*. London: Springer-Verlag, 2005, vol. 53.

[28] T. Fetz and M. Oberguggenberger, "Propagation of uncertainty through multivariate functions in the framework of sets of probability measures," *Reliability Engineering & System Safety*, vol. 85, no. 1–3, pp. 73–87, jul 2004.

[29] T. Fetz and M. Oberguggenberger, "Imprecise random variables, random sets, and monte carlo simulation," *International Journal of Approximate Reasoning*, vol. 78, pp. 252–264, 2016.

[30] Y. Pawitan, *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.

[31] P. Walley and T. L. Fine, "Towards a frequentist theory of upper and lower probability," *The Annals of Statistics*, pp. 741–761, 1982.

[32] M. E. Cattaneo, "Empirical interpretation of imprecise probabilities," in *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, 2017, pp. 61–72.

[33] D. Dubois and H. Prade, "Random sets and fuzzy interval analysis," *Fuzzy Sets and Systems*, vol. 42, no. 1, pp. 87–101, 1991.

[34] D. Dubois and H. Prade,"Interval-valued fuzzy sets, possibility theory and imprecise probability." in *EUSFLAT Conf.*, 2005, pp. 314–319.

[35] D. Dubois and H. Prade, *Possibility theory: an approach to computerized processing of uncertainty*. Springer Science & Business Media, 2012.

[36] R. E. Moore, *Methods and applications of interval analysis*. SIAM, 1979, vol. 2.

[37] K. Weichselberger, "The theory of interval-probability as a unifying concept for uncertainty," *International Journal of Approximate Reasoning*, vol. 24, no. 2, pp. 149–170, 2000.

[38] G. J. Klir, *Uncertainty and information: foundations of generalized information theory*. John Wiley & Sons, 2005.

BIBLIOGRAPHY

[39] Y. Ben-Haim and I. Elishakoff, *Convex models of uncertainty in applied mechanics*. Elsevier, 2013, vol. 25.

[40] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[41] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, pp. 325–339, 1967.

[42] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.

[43] M. Beer, S. Ferson, and V. Kreinovich, "Imprecise probabilities in engineering analyses," *Mechanical systems and signal processing*, vol. 37, no. 1-2, pp. 4–29, 2013.

[44] S. John Walker, "Big data: A revolution that will transform how we live, work, and think," 2014.

[45] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton, "Big data: the management revolution," *Harvard business review*, vol. 90, no. 10, pp. 60–68, 2012.

[46] C. M. Bishop, "Pattern recognition and machine learning," 2006.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

BIBLIOGRAPHY

[48] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[49] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[50] E. F. Halpern, M. C. Weinstein, M. G. M. Hunink, and G. S. Gazelle, "Representing both first-and second-order uncertainties by Monte Carlo simulation for groups of patients," *Medical Decision Making*, vol. 20, no. 3, pp. 314–322, 2000.

[51] J. Zhang and M. D. Shields, "Efficient propagation of imprecise probabilities," *7th International Workshop on Reliable Engineering Computing (REC2016)*, 2016.

[52] M. D. Shields and J. Zhang, "How much data do i really need to conduct probabilistic uq," in *USACM-Workshop on Uncertainty Quantification and Data-Driven Modeling*, 2017.

[53] A. B. Satish, J. Zhang, P. Woelke, and M. Shields, "Probabilistic calibration of material models from limited data and its influence on structural response," in *12th International Conference on Structural Safety and Reliability*, 2017.

BIBLIOGRAPHY

[54] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical science*, pp. 382–401, 1999.

[55] J. M. Bernardo and R. Rueda, "Bayesian hypothesis testing: A reference approach," *International Statistical Review*, vol. 70, no. 3, pp. 351–372, 2002.

[56] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[57] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[58] K. P. Burnham and D. R. Anderson, "Multimodel inference understanding aic and bic in model selection," *Sociological methods & research*, vol. 33, no. 2, pp. 261–304, 2004.

[59] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[60] S. Sankararaman and S. Mahadevan, "Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data," *Reliability Engineering & System Safety*, vol. 96, no. 7, pp. 814–824, 2011.

[61] J. Zhang and M. D. Shields, "The effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets," *Computer Methods in Applied Mechanics and Engineering*, vol. 334, pp. 483–506, 2018.

[62] E. Torre, S. Marelli, P. Embrechts, and B. Sudret, "A general framework for uncertainty quantification under non-gaussian input dependencies," *arXiv preprint arXiv:1709.08626*, 2017.

[63] F. Wang and H. Li, "Subset simulation for non-gaussian dependent random variables given incomplete probability information," *Structural Safety*, vol. 67, pp. 105–115, 2017.

[64] Á. Rózsás and Z. Mogyorósi, "The effect of copulas on time-variant reliability involving time-continuous stochastic processes," *Structural Safety*, vol. 66, pp. 94–105, 2017.

[65] X.-S. Tang, D.-Q. Li, C.-B. Zhou, and K.-K. Phoon, "Copula-based approaches for evaluating slope reliability under incomplete probability information," *Structural Safety*, vol. 52, pp. 90–99, 2015.

[66] W. P. Warsido and G. T. Bitsuamlak, "Synthesis of wind tunnel and climatological data for estimating design wind effects: A copula based approach," *Structural Safety*, vol. 57, pp. 8–17, 2015.

[67] R. Schefzik, T. L. Thorarinsdottir, T. Gneiting *et al.*, "Uncertainty quantification in complex simulation models using ensemble copula coupling," *Statistical science*, vol. 28, no. 4, pp. 616–640, 2013.

[68] Y. Zhang, M. Beer, and S. T. Quek, "Long-term performance assessment and design of offshore structures," *Computers & Structures*, vol. 154, pp. 101–115, 2015.

[69] T. Bedford and R. M. Cooke, "Vines: A new graphical model for dependent random variables," *Annals of Statistics*, pp. 1031–1068, 2002.

[70] H. Joe, *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[71] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[72] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliability Engineering & System Safety*, vol. 91, no. 10-11, pp. 1175–1209, 2006.

[73] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo *et al.*, "Sensitivity

analysis practices: Strategies for model-based inference," *Reliability Engineering & System Safety*, vol. 91, no. 10-11, pp. 1109–1125, 2006.

[74] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 93, no. 7, pp. 964–979, 2008.

[75] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf, "Global sensitivity analysis for systems with independent and/or correlated inputs," *The Journal of Physical Chemistry A*, vol. 114, no. 19, pp. 6022–6032, 2010.

[76] Z. Hu and S. Mahadevan, "Global sensitivity analysis-enhanced surrogate (gsas) modeling for reliability analysis," *Structural and Multidisciplinary Optimization*, vol. 53, no. 3, pp. 501–521, 2016.

[77] Q. Shao, A. Younes, M. Fahs, and T. A. Mara, "Bayesian sparse polynomial chaos expansion for global sensitivity analysis," *Computer Methods in Applied Mechanics and Engineering*, vol. 318, pp. 474–496, 2017.

[78] J.-L. Christen, M. Ichchou, B. Troclet, O. Bareille, and M. Ouisse, "Global sensitivity analysis and uncertainties in sea models of vibroacoustic systems," *Mechanical Systems and Signal Processing*, vol. 90, pp. 365–377, 2017.

[79] M. Oberguggenberger and W. Fellin, "Assessing the sensitivity of failure probabilities: a random set approach," *Safety and Reliability of Engineering Systems and Structures. ICOSSAR*, pp. 1755–1760, 2005.

[80] M. Oberguggenberger, J. King, and B. Schmelzer, "Classical and imprecise probability methods for sensitivity analysis in engineering: A case study," *International Journal of Approximate Reasoning*, vol. 50, no. 4, pp. 680–693, 2009.

[81] J. C. Helton, J. D. Johnson, W. Oberkampf, and C. J. Sallaberry, "Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty," *Reliability Engineering & System Safety*, vol. 91, no. 10-11, pp. 1414–1434, 2006.

[82] C. Li and S. Mahadevan, "Role of calibration, validation, and relevance in multi-level uncertainty integration," *Reliability Engineering & System Safety*, vol. 148, pp. 32–43, 2016.

[83] C. Li and S. Mahadevan,"Relative contributions of aleatory and epistemic uncertainty sources in time series prediction," *International Journal of Fatigue*, vol. 82, pp. 474–486, 2016.

[84] R. Schöbi and B. Sudret, "Uncertainty propagation of p-boxes using sparse polynomial chaos expansions," *Journal of Computational Physics*, vol. 339, pp. 307–327, 2017.

[85] R. Schöbi and B. Sudret, "Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions," *arXiv preprint arXiv:1705.10061*, 2017.

[86] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[87] D. L. Weakliem, "A critique of the bayesian information criterion for model selection," *Sociological Methods & Research*, vol. 27, no. 3, pp. 359–397, 1999.

[88] C. M. Hurvich and C.-L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

[89] C. M. Hurvich and C.-L. Tsai, "Model selection for extended quasi-likelihood models in small samples," *Biometrics*, pp. 1077–1084, 1995.

[90] H. Akaike, "Canonical correlation analysis of time series and the use of an information criterion," *Mathematics in Science and Engineering*, vol. 126, pp. 27–96, 1976.

[91] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[92] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling*. Springer Science & Business Media, 2008.

[93] A. E. Raftery, "Bayesian model selection in social research," *Sociological methodology*, pp. 111–163, 1995.

[94] J. L. Beck and K.-V. Yuen, "Model selection using response measurements: Bayesian probabilistic approach," *Journal of Engineering Mechanics*, vol. 130, no. 2, pp. 192–203, 2004.

[95] J. L. Beck, "Bayesian system identification based on probability logic," *Structural Control and Health Monitoring*, vol. 17, no. 7, pp. 825–847, 2010.

[96] S. H. Cheung and J. L. Beck, "Calculation of posterior probabilities for bayesian model class assessment and averaging from posterior samples based on dynamic system data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 5, pp. 304–321, 2010.

[97] K. Farrell, J. T. Oden, and D. Faghihi, "A bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems," *Journal of Computational Physics*, vol. 295, pp. 189–208, 2015.

[98] J. T. Oden, E. A. Lima, R. C. Almeida, Y. Feng, M. N. Rylander,

D. Fuentes, D. Faghihi, M. M. Rahman, M. DeWitt, M. Gadde *et al.*, "Toward predictive multiscale modeling of vascular tumor growth," *Archives of Computational Methods in Engineering*, vol. 23, no. 4, pp. 735–779, 2016.

[99] E. Prudencio, P. Bauman, D. Faghihi, K. Ravi-Chandar, and J. Oden, "A computational framework for dynamic data-driven material damage control, based on bayesian inference and model selection," *International Journal for Numerical Methods in Engineering*, vol. 102, no. 3-4, pp. 379–403, 2015.

[100] J. Skilling, "Nested sampling," in *AIP Conference Proceedings*, vol. 735, no. 1. AIP, 2004, pp. 395–405.

[101] S. Chib and I. Jeliazkov, "Marginal likelihood from the metropolis–hastings output," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 270–281, 2001.

[102] C. S. Bos, "A comparison of marginal likelihood computation methods," in *Compstat*. Springer, 2002, pp. 111–116.

[103] N. Friel and J. Wyse, "Estimating the evidence–a review," *Statistica Neerlandica*, vol. 66, no. 3, pp. 288–308, 2012.

[104] Z. Zhao and T. A. Severini, "Integrated likelihood computation methods," *Computational Statistics*, pp. 1–33, 2016.

[105] P. Diaconis and D. Ylvisaker, "Conjugate priors for exponential families," *The Annals of statistics*, pp. 269–281, 1979.

[106] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 721–741, 1984.

[107] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[108] J. Goodman and J. Weare, "Ensemble samplers with affine invariance," *Communications in applied mathematics and computational science*, vol. 5, no. 1, pp. 65–80, 2010.

[109] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The mcmc hammer," *Publications of the Astronomical Society of the Pacific*, vol. 125, no. 925, p. 306, 2013.

[110] I. Csisz *et al.*, "Eine informationstheoretische {U} ngleichung und ihre anwendung auf den {B} eweis der ergodizit {\" a} t von {M} arkoffschen {K} etten," *Publ. Math. Inst. Hungar. Acad.*, vol. 8, pp. 95–108, 1963.

[111] T. Morimoto, "Markov processes and the h-theorem," *Journal of the Physical Society of Japan*, vol. 18, no. 3, pp. 328–331, 1963.

[112] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 131–142, 1966.

[113] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen." *Journal für die reine und angewandte Mathematik*, vol. 136, pp. 210–271, 1909.

[114] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[115] R. Beran, "Minimum hellinger distance estimates for parametric models," *The Annals of Statistics*, pp. 445–463, 1977.

[116] B. G. Lindsay, "Efficiency versus robustness: the case for minimum hellinger distance and related methods," *The annals of statistics*, pp. 1081–1114, 1994.

[117] J. Zhang and M. D. Shields, "Probability measure changes in monte carlo simulation," *arXiv preprint arXiv:1803.09121*, 2018.

[118] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal*

of the American Statistical Association, vol. 95, no. 449, pp. 135–143, 2000.

[119] D. Faulkner, "A review of effective plating to be used in the analysis of stiffened plating in bending and compression," Tech. Rep., 1973.

[120] C. A. Carlsen, "Simplified collapse analysis of stiffened plates," *Norwegian Maritime Research*, vol. 5, no. 4, 1977.

[121] J. O. Berger and J. M. Bernardo, "Estimating a product of means: Bayesian analysis with reference priors," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 200–207, 1989.

[122] A. Gelman *et al.*, "Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)," *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[123] L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*. SIAM, 2017, vol. 3.

[124] J. K. Ghosh *et al.*, "Noninformative priors," in *Higher Order Asymptotics*. IMS and ASA, 1994, pp. 86–98.

[125] H. Jeffreys, "An invariant form for the prior probability in estimation problems," in *Proceedings of the Royal Society of London a: mathemat-*

ical, physical and engineering sciences, vol. 186, no. 1007. The Royal Society, 1946, pp. 453–461.

[126] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[127] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[128] H. Chipman, E. I. George, R. E. McCulloch, M. Clyde, D. P. Foster, and R. A. Stine, "The practical implementation of bayesian model selection," *Lecture Notes-Monograph Series*, pp. 65–134, 2001.

[129] A. Mansour, H. Jan, C. Zigelman, Y. Chen, and S. Harding, "Implementation of reliability methods to marine structures," *Transactions-Society of Naval Architects and Marine Engineers*, vol. 92, pp. 353–382, 1984.

[130] K. Atua, I. Assakkaf, and B. M. Ayyub, "Statistical characteristics of strength and load random variables of ship structures," in *Probabilistic Mechanics and Structural Reliability, Proceeding of the Seventh Specialty Conference, Worcester Polytechnic Institute, Worcester, Massachusetts*, 1996.

[131] J. Gabriel and E. Imbembo, "Investigation of the notch-toughness

properties of abs ship platesteels," SHIP STRUCTURE COMMITTEE WASHINGTON DC, Tech. Rep., 1962.

[132] I. BOULGER and W. Hansen, "The effectof metallurgical variables ship-plate steelson the transition temperatures in the drop-weight 1 and charpy v-notch tests," 1962.

[133] J. Kufman and M. Prager, "Marine structural steel toughness data bank. volume 1-4," DTIC Document, Tech. Rep., 1990.

[134] S. Ferson, W. L. Oberkampf, and L. Ginzburg, "Model validation and predictive capability for the thermal challenge problem," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29, pp. 2408–2430, 2008.

[135] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer methods in applied mechanics and engineering*, vol. 200, no. 25, pp. 2131–2144, 2011.

[136] A. Nataf, "Determination des distribution don t les marges sont donnees," *Comptes Rendus de l Academie des Sciences*, vol. 225, pp. 42–43, 1962.

[137] R. Lebrun and A. Dutfoy, "An innovating analysis of the nataf

transformation from the copula viewpoint," *Probabilistic Engineering Mechanics*, vol. 24, no. 3, pp. 312–320, 2009.

[138] M. Sklar, "Fonctions de repartition an dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959.

[139] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[140] E. Brechmann and U. Schepsmeier, "Cdvine: Modeling dependence with c-and d-vine copulas in r," *Journal of Statistical Software*, vol. 52, no. 3, pp. 1–27, 2013.

[141] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 182–198, 2009.

[142] C. Czado, "Pair-copula constructions of multivariate copulas," in *Copula theory and its applications*. Springer, 2010, pp. 93–109.

[143] C. Czado, U. Schepsmeier, and A. Min, "Maximum likelihood estimation of mixed c-vines with application to exchange rates," *Statistical Modelling*, vol. 12, no. 3, pp. 229–255, 2012.

[144] A. Min and C. Czado, "Bayesian model selection for d-vine pair-copula

constructions," *Canadian Journal of Statistics*, vol. 39, no. 2, pp. 239–258, 2011.

[145] J. Dissmann, E. C. Brechmann, C. Czado, and D. Kurowicka, "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics & Data Analysis*, vol. 59, pp. 52–69, 2013.

[146] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[147] J. H. McDonald, *Handbook of biological statistics*. sparky house publishing Baltimore, MD, 2009, vol. 2.

[148] J. Stöber and U. Schepsmeier, "Is there significant time-variation in multivariate copulas?" *arXiv preprint arXiv:1205.4841*, 2012.

[149] L. Gruber, C. Czado *et al.*, "Sequential bayesian model selection of regular vine copulas," *Bayesian Analysis*, vol. 10, no. 4, pp. 937–963, 2015.

[150] R. Younes, A. Hallal, F. Fardoun, and F. H. Chehade, "Comparative review study on elastic properties modeling for unidirectional composite materials," in *Composites and their properties*. intech, 2012.

[151] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, 2010.

[152] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and computers in simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.

[153] L. Le Gratiet, S. Marelli, and B. Sudret, "Metamodel-based sensitivity analysis: Polynomial chaos expansions and gaussian processes," *Handbook of Uncertainty Quantification*, pp. 1289–1325, 2017.

[154] C. Li and S. Mahadevan, "An efficient modularized sample-based method to estimate the first-order sobol index," *Reliability Engineering & System Safety*, vol. 153, pp. 110–121, 2016.

[155] S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz *et al.*, "Estimating the approximation error when fixing unessential factors in global sensitivity analysis," *Reliability Engineering & System Safety*, vol. 92, no. 7, pp. 957–960, 2007.

[156] E. Myshetskaya *et al.*, "Monte carlo estimators for small sensitivity indices," *Monte Carlo Methods and Applications mcma*, vol. 13, no. 5-6, pp. 455–465, 2008.

[157] A. B. Owen, "Better estimation of small sobol'sensitivity indices," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 23, no. 2, p. 11, 2013.

[158] I. M. Sobol, "On quasi-monte carlo integrations," *Mathematics and computers in simulation*, vol. 47, no. 2-5, pp. 103–112, 1998.

[159] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[160] M. D. Shields and J. Zhang, "The generalization of latin hypercube sampling," *Reliability Engineering & System Safety*, vol. 148, pp. 96–108, 2016.

[161] P. Soden, M. Hinton, and A. Kaddour, "Lamina properties, lay-up configurations and loading conditions for a range of fibre reinforced composite laminates," in *Failure Criteria in Fibre-Reinforced-Polymer Composites*. Elsevier, 2004, pp. 30–51.

[162] Z. H. Karadeniz and D. Kumlutas, "A numerical study on the coefficients of thermal expansion of fiber reinforced composite materials," *Composite Structures*, vol. 78, no. 1, pp. 1–10, 2007.

[163] M. K. Chati and A. K. Mitra, "Prediction of elastic properties of

fiber-reinforced unidirectional composites," *Engineering analysis with boundary elements*, vol. 21, no. 3, pp. 235–244, 1998.

[164] Z.-m. Huang, "Micromechanical prediction of ultimate strength of transversely isotropic fibrous composites," *International journal of solids and structures*, vol. 38, no. 22-23, pp. 4147–4172, 2001.

[165] A. Wongsto and S. Li, "Micromechanical fe analysis of ud fibre-reinforced composites with fibres distributed at random over the transverse cross-section," *Composites Part A: Applied Science and Manufacturing*, vol. 36, no. 9, pp. 1246–1266, 2005.

[166] I. M. Daniel, O. Ishai, I. M. Daniel, and I. Daniel, *Engineering mechanics of composite materials*. Oxford university press New York, 1994, vol. 3.

[167] P. K. Mallick, *Composites engineering handbook*. CRC Press, 1997.

[168] A. Letton and W. Bradley, "Studies in long term durability of composites in sea water," in *Proc., Conf. on Use of Composite Mat. in Load-Bearing Marine Structures*, vol. 2, 1990, pp. 163–177.

[169] G. Lubin, *Handbook of composites*. Springer Science & Business Media, 2013.

[170] D. Gay, S. V. Hoa, and S. W. Tsai, *Composite materials: design and applications*. CRC press, 2002.

BIBLIOGRAPHY

[171] E. J. Barbero, *Introduction to composite materials design*. CRC press, 2017.

[172] D. Hull and T. W. Clyne, *An introduction to composite materials*. Cambridge university press, 1996.

[173] H.-Z. Shan and T.-W. Chou, "Transverse elastic moduli of unidirectional fiber composites with fiber/matrix interfacial debonding," *Composites Science and Technology*, vol. 53, no. 4, pp. 383–391, 1995.

[174] A. Pregoretti, M. Traina, and A. Bunsell, "Handbook of tensile properties of textile and technical fibers," 2009.

[175] T. Lamb *et al.*, "Ship design and construction," 2003.

[176] N. P. Cheremisinoff, *Handbook of ceramics and composites*. M. Dekker, 1990.

[177] B. W. Rosen, "Fiber composite materials," *American Society for Metals, Metals Park, Ohio*, vol. 37, 1965.

[178] M. Greyson, "Encyclopedia of composite materials and components, 1983," *Wiley&Sons, USA*.

[179] L. J. Broutman and R. H. Krock, *Modern composite materials*. Addison-Wesley Publishing Company, 1967.

BIBLIOGRAPHY

[180] B. Z. Jang, "Advanced polymer composites: principles and applications," *ASM International, Materials Park, OH 44073-0002, USA, 1994. 305*, 1994.

[181] P. N. Balaguru and S. P. Shah, *Fiber-reinforced cement composites*, 1992.

[182] V. K. Thakur, *Biomass-based biocomposites*. Smithers Rapra, 2013.

[183] R. M. Jones and C. Bert, "Mechanics of composite materials," *Journal of Applied Mechanics*, vol. 42, p. 748, 1975.

[184] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

Vita

Jiaxin Zhang was born in Shenyang, Liaoning, China. He attended Northeast Yucai School from 2001 to 2007. He received his B.S. degree in Engineering Mechanics and M.S. degree in Computational Mechanics from Dalian University of Technology, China, in 2011 and 2014. Jiaxin enrolled in the Civil Engineering at the Johns Hopkins University from August 2014 to August 2018, where he received his M.S.E. degree in Applied Mathematics & Statistics and Ph.D. degree in Civil Engineering. His research focused on uncertainty quantification, stochastic simulation and data-driven modeling, particular for small datasets.

Starting in August 2018, Jiaxin will pursue his academic career at Oak Ridge National Laboratory.